PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 1/21, C07K 14/47, 16/18 | A2

(11) International Publication Number: WO 98/11220

(43) International Publication Date: 19 March 1998 (19.03.98)

(21) International Application Number: PCT/US97/16174

(22) International Filing Date: 12 September 1997 (12.09.97)

(30) Priority Data:
08/712,708 12 September 1996 (12.09.96) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/712,708 (CIP)
Filed on 12 September 1996 (12.09.96)

(71) Applicant (*for all designated States except US*): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). ZWEIGER, Gary, B. [US/US]; 513 S. Fremont Street, San Mateo, CA 94402 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: NOVEL HUMAN CELL DIVISION CYCLE PROTEINS

| Library | Lib Description | Abun | Pct Abun |
|---|---|---|---|
| FIBRNGT01 | GD23A fibroblasts, radiation 5 min | 1 | 0.1664 |
| PITUNOR01 | pituitary, 16-70 M/F, RP | 1 | 0.1233 |
| MYOMNOT01 | uterus, myometrium, 43 F | 1 | 0.0409 |
| STOMTUT01 | stomach tumor, 52 M, match to STOMNOT02 | 1 | 0.0367 |
| BRAITUT02 | brain tumor, metastasis, 58 M | 2 | 0.0338 |
| STOMNOT02 | stomach, 52 M, match to STOMTUT01 | 1 | 0.0308 |
| LUNGNOT09 | lung, fetal M | 1 | 0.0286 |
| PTHYTUM01 | parathyroid tumor, adenoma, M/F, NORM, WM | 1 | 0.0278 |
| LNODNOT03 | lymph node, 67 M | 1 | 0.0265 |
| BRAITUT13 | brain tumor, meningioma, 68 M | 1 | 0.0262 |
| DUODNOT02 | small intestine, duodenum, 8 F | 1 | 0.0262 |
| BRAINOT03 | brain, 26 M | 1 | 0.0185 |
| HNT2RAT01 | hNT-2 cell line, teratocarcinoma, treated RA | 1 | 0.0185 |
| LUNGNOT04 | lung, 2 M | 1 | 0.0183 |
| UTRSNOT02 | uterus, 34 F | 1 | 0.0166 |
| NGANNOT01 | ganglioneuroma, 9 M | 1 | 0.0155 |
| BRAINOM01 | brain, infant F, NORM, WM | 3 | 0.0134 |
| UCMCL5T01 | mononuclear cells, treated IL-5 | 1 | 0.0125 |

(57) Abstract

The present invention provides novel human cell division cycle proteins (collectively called HCDC) and polynucleotides which identify and encode HCDC. The invention also provides genetically engineered expression vectors and host cells comprising the nucleic acid sequences encoding HCDC. The invention also provides pharmaceutical compositions containing HCDC or antagonists to HCDC, and in the use of these compositions for the treatment of diseases associated with the expression of HCDC. Additionally, the invention provides for the use of antisense molecules to polynucleotides encoding HCDC for the treatment of diseases associated with the expression of HCDC. The invention also provides diagnostic assays which utilize the polynucleotide, or fragments or the complement thereof, to hybridize to the genomic sequence or transcripts of polynucleotides encoding HCDC or anti-HCDC antibodies which specifically bind to HCDC.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AL | Albania |
| AM | Armenia |
| AT | Austria |
| AU | Australia |
| AZ | Azerbaijan |
| BA | Bosnia and Herzegovina |
| BB | Barbados |
| BE | Belgium |
| BF | Burkina Faso |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| BY | Belarus |
| CA | Canada |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CI | Côte d'Ivoire |
| CM | Cameroon |
| CN | China |
| CU | Cuba |
| CZ | Czech Republic |
| DE | Germany |
| DK | Denmark |
| EE | Estonia |
| ES | Spain |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| GE | Georgia |
| GH | Ghana |
| GN | Guinea |
| GR | Greece |
| HU | Hungary |
| IE | Ireland |
| IL | Israel |
| IS | Iceland |
| IT | Italy |
| JP | Japan |
| KE | Kenya |
| KG | Kyrgyzstan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| KZ | Kazakstan |
| LC | Saint Lucia |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LR | Liberia |
| LS | Lesotho |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| MC | Monaco |
| MD | Republic of Moldova |
| MG | Madagascar |
| MK | The former Yugoslav Republic of Macedonia |
| ML | Mali |
| MN | Mongolia |
| MR | Mauritania |
| MW | Malawi |
| MX | Mexico |
| NE | Niger |
| NL | Netherlands |
| NO | Norway |
| NZ | New Zealand |
| PL | Poland |
| PT | Portugal |
| RO | Romania |
| RU | Russian Federation |
| SD | Sudan |
| SE | Sweden |
| SG | Singapore |
| SI | Slovenia |
| SK | Slovakia |
| SN | Senegal |
| SZ | Swaziland |
| TD | Chad |
| TG | Togo |
| TJ | Tajikistan |
| TM | Turkmenistan |
| TR | Turkey |
| TT | Trinidad and Tobago |
| UA | Ukraine |
| UG | Uganda |
| US | United States of America |
| UZ | Uzbekistan |
| VN | Viet Nam |
| YU | Yugoslavia |
| ZW | Zimbabwe |

# NOVEL HUMAN CELL DIVISION CYCLE PROTEINS

## TECHNICAL FIELD

The present invention relates to nucleic acid and amino acid sequences of novel human cell division cycle proteins and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

## BACKGROUND ART

Much has been learned about the process of cyclical growth and division of eukaryotic cells through the identification and characterization of cell division cycle (cdc) mutants in budding yeast. Cdc36 and Cdc37 are among several temperature-sensitive mutants which arrest in the G1 phase of the yeast Saccharomyces cerevisiae cell cycle (Shuster JR (1982) Mol Cell Biol 2:1052-1063; Reed SI (1980) Genetics 95 561-577). The yeast genes CDC36 and CDC37 were identified by complementation of the respective yeast mutant, cloned and sequenced (Breter HJ et al (1983) Mol Cell Biol 3:881-891; Ferguson J et al (1986) Nucleic Acids Res 14:6681-6697).

CDC36 (also referred to as NOT2) was one of several yeast genes discovered in a search for genes that preferentially affect and negatively regulate transcription that depends upon the $T_C$ TATA basal level transcription element (Collart MA et al (1994) Genes and Devel 8:525-537). Cdc36 is part of a 500 kD nucleus localized complex and is likely to inhibit the basic RNA polymerase II transcription machinery necessary for cell cycle progression, as well as many other important cell processes (Collart et al, supra). Cdc36 has homology to a portion of an oncogenic protein, the ets product from the avian erythroblastosis virus E26 (Peterson TA et al (1984) Nature 309:556-558) and an open reading frame (ORF; GI 1053220) of a C. elegans cDNA (Wilson R et al (1994) Nature 368:32-38). No vertebrate Cdc36 homologs have been reported.

Cdc37, however, has homology to avian (Grammatikakis N et al (1995) J Biol Chem 270: 16198-16205) and mammalian (Stepanova L et al (1996) Genes and Devel 10:1491-1502) sequences. In fact Cdc37 is identical to mammalian p50, a protein known to interact with the oncogenes pp60$^{v\text{-}src}$ and Raf-1 (Stepanova et al, supra). Experiments with mouse fibroblasts and insect cells showed that Cdc37 forms a complex with the chaperone protein Hsp90 and helps stabilize Cdk4, a kinase with an important role in progression through the G1 phase of the cell cycle (Stepanova, supra).

### Cell Division Cycle and Disease

Progression through the cell cycle, and consequently cell proliferation, are governed by

the complex interactions of protein complexes composed of cyclins, cyclin-dependent protein kinases, and associated proteins (Cordon-Cardo C (1995) Am J Pathol 147:545-560). Cancers are characterized by uncoordinated cell proliferation and can be identified by changes in the protein complexes that normally control progression through the cell cycle (Nigg EA (1995) Bioessays 17:471-480). A primary treatment for cancer involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle control (Neubauer A et al (1996) Leukemia 10:S2-S4). For example, Cordon-Cardo (supra) suggested that negative regulators of Cdk4 may act as tumor suppressors.

Experiments with breast cancer and erythroleukemia cells show that certain agents which halt cell growth are probably acting through an inhibition of Cdk4 activity (Watts CK et al (1995) Mol Endocrinol 9:1804-1813; Marks PA et al (1994) Proc Natl Acad Sci 91:10251-10254). The TATA box-dependent transcription machinery is also a potential target for cancer therapeutics. An analogous situation is demonstrated with the tumor suppressor protein p53, which represses the activity of promoters whose initiation is dependent on the presence of a TATA box (Mack DH et al (1993) Nature 363: 81-283). Furthermore, Mack et al (supra) observed that p53 repression is mediated by an interaction of p53 with basal transcription factors.

Modulation of factors which act in the coordination of the human cell division cycle may provide an important means by which to stop cancer cell growth. Thus, new cell division cycle proteins which modulate these processes could satisfy a significant need in the art by providing new means of diagnosing and treating cancer.

## DISCLOSURE OF THE INVENTION

The present invention discloses two novel human cell division cycle proteins (hereinafter referred to individually as HCDCA and HCDCB, and collectively as HCDC), characterized as having homology to avian Cdc37 (GI 755484) and yeast Cdc36 (GI 115930), respectively. Accordingly, the invention features two substantially purified cell division cycle proteins, having the amino acid sequence shown in SEQ ID NO:1 and SEQ ID NO:3, and having characteristics of cell division cycle proteins.

One aspect of the invention features isolated and substantially purified polynucleotides which encode HCDC. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:2 or SEQ ID NO:4. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:2 or SEQ ID NO:4.

The invention further relates to nucleic acid sequences encoding HCDC, oligonucleotides,

peptide nucleic acids (PNA), fragments, portions or antisense molecules thereof, and expression vectors and host cells comprising polynucleotides which encode HCDC. The present invention also relates to antibodies which bind specifically to HCDC and pharmaceutical compositions comprising substantially purified HCDC or fragments thereof, or antagonists of HCDC, and methods for producing HCDC or fragments thereof.

## BRIEF DESCRIPTION OF DRAWINGS

Figures 1A, 1B, 1C and 1D show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of the novel cell division cycle protein, HCDCA. The alignment was produced using MacDNAsis software (Hitachi Software Engineering Co Ltd, San Bruno, CA).

Figures 2A, 2B, 2C and 2D show the amino acid sequence (SEQ ID NO:3) and nucleic acid sequence (SEQ ID NO:4) of the novel cell division cycle protein, HCDCB (MacDNAsis software, Hitachi Software Engineering Co Ltd).

Figures 3A, 3B, 3C and 3D show the northern analysis for SEQ ID NO:2. The northern analysis was produced electronically using LIFESEQ™ database (Incyte Pharmaceuticals, Palo Alto CA).

Figure 4 shows the northern analysis for SEQ ID NO:4 (LIFESEQ™ database, Incyte Pharmaceuticals, Palo Alto CA).

Figures 5A, 5B and 5C show the amino acid sequence alignments among HCDCA (SEQ ID NO:1), avian Cdc37 (GI 755484; SEQ ID NO:5), rat Cdc37 (GI 1197180; SEQ ID NO:6), and yeast Cdc37 (GI 1077057; SEQ ID NO:7) produced using the multisequence alignment program of DNAStar software (DNAStar Inc, Madison WI).

Figures 6A and 6B shows the amino acid sequence alignments among HCDCB (SEQ ID NO:3), an ORF of C. elegans cDNA (GI 1053220; SEQ ID NO:8), and yeast Cdc36 (GI 115930; SEQ ID NO:9), produced using the multisequence alignment program of DNAStar software (DNAStar Inc, Madison WI).

Figure 7 shows the hydrophobicity plot (generated using MacDNAsis software) for HCDCA, SEQ ID NO:1; the X axis reflects amino acid position, and the negative Y axis, hydrophobicity (Figs. 7, 8, 9, and 10).

Figure 8 shows the hydrophobicity plot for rat Cdc37, SEQ ID NO:6.

Figure 9 shows the hydrophobicity plot for HCDCB, SEQ ID NO:3.

Figure 10 shows the hydrophobicity plot for yeast Cdc36, SEQ ID NO:9.

## MODES FOR CARRYING OUT THE INVENTION

**Definitions**

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. Similarly, amino acid sequence as used herein refers to peptide or protein sequence.

"Peptide nucleic acid" as used herein refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen PE et al (1993) Anticancer Drug Des 8:53-63).

As used herein, HCDC refers to the amino acid sequences of substantially purified HCDC obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic or recombinant.

"Consensus" as used herein may refer to a nucleic acid sequence 1) which has been resequenced to resolve uncalled bases, 2) which has been extended using XL-PCR (Perkin Elmer) in the 5' or the 3' direction and resequenced, 3) which has been assembled from the overlapping sequences of more than one Incyte clone GCG Fragment Assembly System, (GCG, Madison WI), or 4) which has been both extended and assembled.

A "variant" of HCDC is defined as an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, eg, replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, eg, replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNAStar software.

A "deletion" is defined as a change in either amino acid or nucleotide sequence in which one or more amino acid or nucleotide residues, respectively, are absent.

An "insertion" or "addition" is that change in an amino acid or nucleotide sequence which has resulted in the addition of one or more amino acid or nucleotide residues, respectively, as

compared to the naturally occurring HCDC.

A "substitution" results from the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

The term "biologically active" refers to an HCDC having structural, regulatory or biochemical functions of a naturally occurring HCDC. Likewise, "immunologically active" defines the capability of the natural, recombinant or synthetic HCDC, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "derivative" as used herein refers to the chemical modification of a nucleic acid encoding HCDC or the encoded HCDC. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of natural HCDC.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

"Stringency" typically occurs in a range from about Tm-5°C (5°C below the Tm of the probe)to about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a stringency hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

The term "hybridization" as used herein shall include "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY). Amplification as carried out in the polymerase chain reaction technologies is described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

**Preferred Embodiments**

The present invention relates to novel HCDC and to the use of the nucleic acid and amino acid sequences in the study, diagnosis, prevention and treatment of disease. cDNAs encoding a portion of HCDC were found in cDNA libraries derived from a variety of tissues, including many types of tumors (Figures 3A, 3B, 3C, 3D and 4).

The present invention also encompasses HCDC variants. A preferred HCDC variant is one having at least 90% amino acid sequence similarity to the HCDC amino acid sequence (SEQ

ID NO:1; SEQ ID NO:3) and a most preferred HCDC variant is one having at least 95% amino acid sequence similarity to SEQ ID NO:1 or SEQ ID NO:3.

Nucleic acids encoding the human HCDC of the present invention were first identified in cDNA, Incyte Clones 532234 (brain cDNA library, BRAINOT03) and 613725 (colon tumor cDNA library, COLNTUT02), through a computer-generated search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 532234 (from cDNA library BRAINOT03); 012498 (THP1PLB01); 176292 (TLYMNOT01); 193713 (KIDNNOT02) 222235 (PANCNOT01); 303291 and 304386 (TESTNOT04); 483523 (HNT2RAT01); 490688 (HNT2AGT01); 547705 and 547889 (BEPINOT01); 552573 (SCORNOT01); 587425 (UTRSNOT01); 604958 (BRSTTUT01); 619618 and 622323 (PGANNOT01); 677158 (CRBLNOT0); 724095 and 726301 (SYNOOAT01); 730945 (LUNGNOT03); 751709 (BRAITUT01); 764129, 765754, and 768117 (LUNGNOT04); 818552, 820214, and 822359 (KERANOT02); 834047 and 835535 (PROSNOT07); 903593 (COLNNOT07); 908316 (COLNNOT09); 961898 (BRSTTUT03); 1284032 (COLNNOT16); 1289033 (BRAINOT11); and 1238055 (LUNGTUT02). A consensus sequence, SEQ ID NO:4, was derived from the extended nucleic acid sequence of Incyte Clones 613725 (from cDNA library COLNTUT02).

The HCDCA amino acid sequence, SEQ ID NO:1, is encoded by the nucleic acid sequence of SEQ ID NO:2. SEQ ID NO:1 and SEQ ID NO:2 precisely matches the respective amino acid and nucleotide sequences of human $p50^{Cdc37}$ (Stepanova et al, supra). HCDCB amino acid sequence, SEQ ID NO:3, is encoded by the nucleic acid sequence of SEQ ID NO:4. The present invention is based, in part, on the chemical and structural homology among HCDCA, avian Cdc37 (GI 755484; Grammatikakis et al, supra), rat Cdc37 (GI 1197180; Ozaki et al, supra), and yeast Cdc37 (GI 1077057; Ferguson et al, supra); Figures 5A, 5B and 5C) and among HCDCB, an ORF on C. elegans cDNA (GI 1053220; Wilson et al, supra), and yeast Cdc36 (GI 115930; Ferguson et al 1995, supra; Figures 6A and 6B). HCDCA and avian Cdc37 share 88% identity, whereas HCDCB and yeast Cdc36 share 28% identity. As illustrated by Figures 7-10, HCDCA and rat Cdc37, and HCDCB and yeast Cdc36 have similar hydrophobicity plots suggesting similar structure. The novel HCDCA is 378 amino acids long and the novel HCDCB is 280 amino acids long.

**The HCDC Coding Sequences**

The nucleic acid and deduced amino acid sequences of HCDCA and HCDCB are shown

in Figures 1A, 1B, 1C, 1D, 2A, 2B, 2C and 2D. In accordance with the invention, any nucleic acid sequence which encodes the amino acid sequence of HCDC can be used to generate recombinant molecules which express HCDC. In a specific embodiment described herein, a nucleotide sequence encoding a portion of HCDCA was first isolated as Incyte Clones 532234 from a brain cDNA library (BRAINOT03). In another specific embodiment described herein, a nucleotide sequence encoding a portion of HCDCB was first isolated as Incyte Clones 613725 from a colon tumor cDNA library (COLNTUT02).

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of HCDC-encoding nucleotide sequences, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene may be produced. The invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring HCDC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HCDC and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HCDC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HCDC or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HCDC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

It is now possible to produce a DNA sequence, or portions thereof, encoding an HCDC and its derivatives entirely by synthetic chemistry, after which the synthetic gene may be inserted into any of the many available DNA vectors and cell systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HCDC or any portion thereof.

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridizing to the nucleotide sequences of Figures 1A, 1B, 1C, 1D, 2A, 2B, 2C

and 2D under various conditions of stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and confer may be used at a defined stringency.

Altered nucleic acid sequences encoding HCDC which may be used in accordance with the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent HCDC. The protein may also show deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HCDC. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of HCDC is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine phenylalanine, and tyrosine.

Included within the scope of the present invention are alleles of HCDC. As used herein, an "allele" or "allelic sequence" is an alternative form of HCDC. Alleles result from a mutation, ie, a change in the nucleic acid sequence, and generally produce altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions or substitutions of amino acids. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

Methods for DNA sequencing are well known in the art and employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland OH)), Taq polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the ABI 377 DNA sequencers (Perkin Elmer).

**Extending the P lynucleotide Sequence**

The polynucleotide sequence encoding HCDC may be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. Gobinda et al (1993; PCR Methods Applic 2:318-22) disclose "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia T et al (1988) Nucleic Acids Res 16:8186). The primers may be designed using OLIGO® 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom M et al (1991) PCR Methods Applic 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method which may be used to retrieve unknown sequences is that of Parker JD et al (1991; Nucleic Acids Res 19:3055-60). Additionally, one can use PCR, nested primers and PromoterFinder libraries to walk in genomic DNA (PromoterFinder™ Clontech (Palo Alto CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

Capillary electrophoresis may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. Systems for rapid sequencing are available from Perkin Elmer, Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity is converted to electrical signal using appropriate software (eg. Genotyper™ and Sequence Navigator™ from Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis is particularly suited to the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample. The reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min has been reported (Ruiz-Martinez MC et al (1993) Anal Chem 65:2851-2858).

**Expression of the Nucleotide Sequence**

In accordance with the present invention, polynucleotide sequences which encode HCDC, fragments of the polypeptide, fusion proteins or functional equivalents thereof may be used in recombinant DNA molecules that direct the expression of HCDC in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence, may be used to clone and express HCDC. As will be understood by those of skill in the art, it may be advantageous to produce HCDC-encoding nucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular prokaryotic or eukaryotic host (Murray E et al (1989) Nuc Acids Res 17:477-508) can be selected, for example, to increase the rate of HCDC expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered in order to alter an HCDC coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, eg, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

In another embodiment of the invention, a natural, modified or recombinant polynucleotides encoding HCDC may be ligated to a heterologous sequence to encode a fusion

protein. For example, for screening of peptide libraries for inhibitors of HCDC activity, it may be useful to encode a chimeric HCDC protein that is recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between an HCDC sequence and the heterologous protein sequence, so that the HCDC may be cleaved and purified away from the heterologous moiety.

In an alternate embodiment of the invention, the coding sequence of HCDC may be synthesized, whole or in part, using chemical methods well known in the art (see Caruthers MH et al (1980) Nuc Acids Res Symp Ser 215-23, Horn T et al(1980) Nuc Acids Res Symp Ser 225-32, etc). Alternatively, the protein itself could be produced using chemical methods to synthesize an HCDC amino acid sequence, whole or in part. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge JY et al (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The newly synthesized peptide can be substantially by preparative high performance liquid chromatography (eg, Creighton (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co, New York NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (eg, the Edman degradation procedure; Creighton, supra). Additionally the amino acid sequence of HCDC, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

**Expression Systems**

In order to express a biologically active HCDC, the nucleotide sequence encoding HCDC or its functional equivalent, is inserted into an appropriate expression vector, ie, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing an HCDC coding sequence and appropriate transcriptional or translational controls. These methods include in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination or genetic recombination. Such techniques are described in Sambrook et al (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY and Ausubel FM et al (1989) Current Protocols in Molecular Biology, John Wiley

& Sons, New York NY.

A variety of expression vector/host systems may be utilized to contain and express an HCDC coding sequence. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (eg, baculovirus); plant cell systems transfected with virus expression vectors (eg, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (eg, Ti or pBR322 plasmid); or animal cell systems.

The "control elements" or "regulatory sequences" of these systems vary in their strength and specificities and are those nontranslated regions of the vector, enhancers, promoters, and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla CA) or pSport1 (Gibco BRL) and ptrp-lac hybrids and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (eg, heat shock, RUBISCO; and storage protein genes) or from plant viruses (eg, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from the mammalian genes or from mammalian viruses are most appropriate. If it is necessary to generate a cell line that contains multiple copies of HCDC, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for HCDC. For example, when large quantities of HCDC are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the HCDC coding sequence may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster (1989) J Biol Chem 264:5503-5509); and the like. pGEX vectors (Promega, Madison WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble

and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems are designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel et al (supra) and Grant et al (1987) Methods in Enzymology 153:516-544.

In cases where plant expression vectors are used, the expression of a sequence encoding HCDC may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al (1984) Nature 310:511-514) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al (1987) EMBO J 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al (1984) EMBO J 3:1671-1680; Broglie et al (1984) Science 224:838-843); or heat shock promoters (Winter J and Sinibaldi RM (1991) Results Probl Cell Differ 17:85-105) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S or Murry LE in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill New York NY, pp 191-196 or Weissbach and Weissbach (1988) Methods for Plant Molecular Biology, Academic Press, New York NY, pp 421-463.

An alternative expression system which could be used to express HCDC is an insect system. In one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The HCDC coding sequence may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of HCDC will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect S. frugiperda cells or Trichoplusia larvae in which HCDC is expressed (Smith et al (1983) J Virol 46:584; Engelhard EK et al (1994) Proc Nat Acad Sci 91:3224-7).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, an HCDC coding sequence may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will

result in a viable virus capable of expressing HCDC in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci 81:3655-59). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of an HCDC sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where HCDC, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf D et al (1994) Results Probl Cell Differ 20:125-62; Bittner et al (1987) Methods in Enzymol 153:516-544).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, 293, WI38, etc have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express HCDC may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These

include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I et al (1980) Cell 22:817-23) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler M et al (1980) Proc Natl Acad Sci 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin F et al (1981) J Mol Biol 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes CA et al (1995) Methods Mol Biol 55:121-131).

**Identification of Transformants Containing the Polynucleotide Sequence**

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the HCDC is inserted within a marker gene sequence, recombinant cells containing HCDC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with an HCDC sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem HCDC as well.

Alternatively, host cells which contain the coding sequence for HCDC and express HCDC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of the nucleic acid or protein.

The presence of the polynucleotide sequence encoding HCDC can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of polynucleotides encoding HCDC. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the HCDC-encoding sequence to detect transformants

containing DNA or RNA encoding HCDC. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides which can be used as a probe or amplimer. A variety of protocols for detecting and measuring the expression of HCDC, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HCDC is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton R et al (1990, Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox DE et al (1983, J Exp Med 158:1211).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HCDC include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the HCDC sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced as shown in US Patent No. 4,816,567 incorporated herein by reference.

**Purification of HCDC**

Host cells transformed with a nucleotide sequence encoding HCDC may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the

art, expression vectors containing polynucleotides encoding HCDC can be designed with signal sequences which direct secretion of HCDC through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may join HCDC to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins (Kroll DJ et al (1993) DNA Cell Biol 12:441-53; cf discussion of vectors infra containing fusion proteins).

HCDC may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and HCDC is useful to facilitate purification. One such expression vector provides for expression of a fusion protein compromising an HCDC and contains nucleic acid encoding 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromotography as described in Porath et al (1992) Protein Expression and Purification 3: 263-281) while the enterokinase cleavage site provides a means for purifying HCDC from the fusion protein.

In addition to recombinant production, fragments of HCDC may be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J (1963) J Am Chem Soc 85:2149-2154). In vitro protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance with the instructions provided by the manufacturer. Various fragments of HCDC may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

**Uses of HCDC and Polynucleotides Encoding HCDC**

The rationale for use of the nucleotide and polypeptide sequences disclosed herein is based in part on the chemical and structural homology among the novel HCDCA protein disclosed herein, avian Cdc37 (GI 755484; Grammatikakis et al, supra), rat Cdc37 (GI 1197180; Ozaki et al, supra), and yeast Cdc37 (GI 1077057; Ferguson et al, supra) and among the novel HCDCB, an ORF on C. elegans cDNA (GI 1053220; Wilson et al, supra), and yeast Cdc36 (GI

115930; Ferguson et al, supra). In addition, northern analysis disclosed herein indicates that HCDC molecules are expressed in cells derived from many types of human cancers (Figures 2A, 2B, 2C and 2D).

Both HCDC proteins appear to function in the cell division cycle. Accordingly, HCDC or an HCDC derivative may be used to modulate the cell division cycle, which is integral to the development and spread of cancerous cells. An HCDC protein that acts as a basal transcription factor may promote cancer cell growth. In conditions where HCDC protein activity is not desirable, cells could be transfected with antisense sequences to HCDC-encoding polynucleotides or provided with antagonists to HCDC. Thus, HCDC antagonists or antisense molecules may be used to slow, stop, or reverse cancer cell growth.

**HCDC Antibodies**

HCDC-specific antibodies are useful for the diagnosis of conditions and diseases associated with expression of HCDC. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Neutralizing antibodies, ie, those which inhibit dimer formation, are especially preferred for diagnostics and therapeutics.

HCDC for antibody induction does not require biological activity; however, the protein fragment, or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably, they should mimic a portion of the amino acid sequence of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HCDC amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Procedures well known in the art can be used for the production of antibodies to HCDC.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with HCDC or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal antibodies to HCDC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (1975 Nature 256:495-497), the human B-cell hybridoma technique (Kosbor et al (1983) Immunol Today 4:72; Cote et al (1983) Proc Natl Acad Sci 80:2026-2030) and the EBV-hybridoma technique (Cole et al (1985) Monoclonal Antibodies and Cancer Therapy, Alan R Liss Inc, New York NY, pp 77-96).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison et al (1984) Proc Natl Acad Sci 81:6851-6855; Neuberger et al (1984) Nature 312:604-608; Takeda et al (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies (US Patent No. 4,946,778) can be adapted to produce HCDC-specific single chain antibodies

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi et al (1989, Proc Natl Acad Sci 86:3833-3837), and Winter G and Milstein C (1991; Nature 349:293-299).

Antibody fragments which contain specific binding sites for HCDC may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD et al (1989) Science 256:1275-1281).

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the formation of complexes between HCDC and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on a specific HCDC protein is preferred, but a competitive binding assay may also be employed. These assays are described in Maddox DE et al (1983, J Exp Med 158:1211).

**Diagnostic Assays Using HCDC Specific Antib dies**

Particular HCDC antibodies are useful for the diagnosis of conditions or diseases

characterized by expression of HCDC or in assays to monitor patients being treated with HCDC, agonists or inhibitors. Diagnostic assays for HCDC include methods utilizing the antibody and a label to detect HCDC in human body fluids or extracts of cells or tissues. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, the polypeptides and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known, several of which were described above.

A variety of protocols for measuring HCDC, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HCDC is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, DE et al (1983, J Exp Med 158:1211).

In order to provide a basis for diagnosis, normal or standard values for HCDC expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with antibody to HCDC under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified by comparing various artificial membranes containing known quantities of HCDC with both control and disease samples from biopsied tissues. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

**Drug Screening**

HCDC, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The fragment employed in such a test may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between HCDC and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the HCDC is described in detail in "Determination of Amino Acid Sequence Antigenicity" by Geysen HN, WO Application

84/03564, published on September 13, 1984, and incorporated herein by reference. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with fragments of HCDC and washed. Bound HCDC is then detected by methods well known in the art. Purified HCDC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding HCDC specifically compete with a test compound for binding HCDC. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HCDC.

**Uses of the Polynucleotide Encoding HCDC**

A polynucleotide encoding HCDC, or any part thereof, may be used for diagnostic and/or therapeutic purposes. For diagnostic purposes, polynucleotides encoding HCDC of this invention may be used to detect and quantitate gene expression in biopsied tissues in which expression of HCDC may be implicated. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of HCDC and to monitor regulation of HCDC levels during therapeutic intervention. Included in the scope of the invention are oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs.

Another aspect of the subject invention is to provide for hybridization or PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HCDC or closely related molecules. The specificity of the probe, whether it is made from a highly specific region, eg, 10 unique nucleotides in the 5' regulatory region, or a less specific region, eg, especially in the 3' region, and the stringency of the hybridization or amplification (maximal, high, intermediate or low) will determine whether the probe identifies only naturally occurring sequences encoding HCDC, alleles or related sequences.

Probes may also be used for the detection of related sequences and should preferably contain at least 50% of the nucleotides from any of these HCDC encoding sequences. The hybridization probes of the subject invention may be derived from the nucleotide sequence of SEQ ID NO:2 or from genomic sequence including promoter, enhancer elements and introns of the naturally occurring HCDC. Hybridization probes may be labeled by a variety of reporter groups, including radionuclides such as 32P or 35S, or enzymatic labels such as alkaline

phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Other means for producing specific hybridization probes for DNAs encoding HCDC include the cloning of nucleic acid sequences encoding HCDC or HCDC derivatives into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides.

Polynucleotide sequences encoding HCDC may be used for the diagnosis of conditions or diseases with which the expression of HCDC is associated. For example, polynucleotide sequences encoding HCDC may be used in hybridization or PCR assays of fluids or tissues from biopsies to detect HCDC expression. The form of such qualitative or quantitative methods may include Southern or northern analysis, dot blot or other membrane-based technologies; PCR technologies; dip stick, pIN, chip and ELISA technologies. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits.

The nucleotide sequences encoding HCDC disclosed herein provide the basis for assays that detect activation or induction associated with various cancers. The nucleotide sequence encoding HCDC may be labeled by methods known in the art and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After an incubation period, the sample is washed with a compatible fluid which optionally contains a dye (or other label requiring a developer) if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye in the biopsied or extracted sample is significantly elevated over that of a comparable control sample, the nucleotide sequence has hybridized with nucleotide sequences in the sample, and the presence of elevated levels of nucleotide sequences encoding HCDC in the sample indicates the presence of the associated disease.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regime in animal studies, in clinical trials, or in monitoring the treatment of an individual patient. In order to provide a basis for the diagnosis of disease, a normal or standard profile for HCDC expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with HCDC, or a portion thereof, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained for normal subjects with a dilution series of HCDC run in the

same experiment where a known amount of a substantially purified HCDC is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients afflicted with HCDC-associated diseases. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established, a therapeutic agent is administered and a treatment profile is generated. Such assays may be repeated on a regular basis to evaluate whether the values in the profile progress toward or return to the normal or standard pattern. Successive treatment profiles may be used to show the efficacy of treatment over a period of several days or several months.

PCR, as described in US Patent Nos. 4,683,195 and 4,965,188, provides additional uses for oligonucleotides based upon the HCDC sequence. Such oligomers are generally chemically synthesized, but they may be generated enzymatically or produced from a recombinant source. Oligomers generally comprise two nucleotide sequences, one with sense orientation (5'->3') and one with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Additionally, methods which may be used to quantitate the expression of a particular molecule include radiolabeling (Melby PC et al 1993 J Immunol Methods 159:235-44) or biotinylating (Duplaa C et al 1993 Anal Biochem 229-36) nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated. Quantitation of multiple samples may be speeded up by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation. For example, the presence of a relatively high amount of HCDC in extracts of biopsied tissues may indicate the onset of various cancers. A definitive diagnosis of this type may allow health professionals to begin aggressive treatment and prevent further worsening of the condition. Similarly, further assays can be used to monitor the progress of a patient during treatment. Furthermore, the nucleotide sequences disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, and the like.

**Therapeutic Use**

Based upon its homology to genes encoding cell division cycle proteins and its expression

profile, polynucleotide sequences encoding HCDC disclosed herein may be useful in the treatment of conditions such as cancer.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express antisense polynucleotides of the gene encoding HCDC. See, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra).

The polynucleotides comprising full length cDNA sequence and/or its regulatory elements enable researchers to use sequences encoding HCDC as an investigative tool in sense (Youssoufian H and HF Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991) Annu Rev Biochem 60:631-652) regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Genes encoding HCDC can be turned off by transfecting a cell or tissue with expression vectors which express high levels of a desired HCDC-encoding fragment. Such constructs can flood cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector (Mettler I, personal communication) and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the control regions of gene encoding HCDC, ie, the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, eg, between -10 and +10 regions of the leader sequence, are preferred. The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA were reviewed by Gee JE et al (In: Huber BE and BI Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco NY).

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of

RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HCDC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HCDC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Methods for introducing vectors into cells or tissues include those methods discussed infra and which are equally suitable for in vivo, in vitro and ex vivo therapy. For ex vivo therapy, vectors are introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient is presented in US Patent Nos. 5,399,493 and 5,437,994, disclosed herein by reference. Delivery by transfection and by liposome are quite

well known in the art.

Furthermore, the nucleotide sequences for HCDC disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

**Detection and Mapping of Related Polynucleotide Sequences**

The nucleic acid sequence for HCDC can also be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include in situ hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price CM (1993; Blood Rev 7:127-34) and Trask BJ (1991; Trends Genet 7:149-54).

The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY. Fluorescent in situ hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding HCDC on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. For example an sequence tagged site based map of the human genome was recently published by the Whitehead-MIT Center for Genomic Research (Hudson TJ et al (1995) Science 270:1945-1954). Often the placement of a gene on the chromosome of another mammalian species such as mouse (Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Database Release 10, April 28, 1995) may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned

to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once a disease or syndrome, such as ataxia telangiectasia (AT), has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti et al (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

**Pharmaceutical Compositions**

The present invention relates to pharmaceutical compositions which may comprise nucleotides, proteins, antibodies, agonists, antagonists, or inhibitors, alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone, or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with excipient(s) or pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

**Administration of Pharmaceutical Compositions**

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of "Remington's Pharmaceutical Sciences" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, ie, dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be

permeated are used in the formulation. Such penetrants are generally known in the art.

**Manufacture and Storage**

The pharmaceutical compositions of the present invention may be manufactured in a manner that known in the art, eg, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5 that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HCDC, such labeling would include amount, frequency and method of administration.

**Therapeutically Effective Dose**

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, eg, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, eg, ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account include the severity of the disease state, eg, tumor size and location; age, weight and gender of the patient; diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long acting pharmaceutical compositions might be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

It is contemplated, for example, that HCDC or an HCDC derivative can be delivered in a suitable formulation to block the progression of various cancers. Similarly, administration of HCDC antagonists may also inhibit the activity or shorten the lifespan of this protein.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

## INDUSTRIAL APPLICABILITY

### I Construction of cDNA Libraries

<u>Colon Tumor</u>

The COLNTUT02 cDNA library was constructed from tissue of a colon tumor removed from a 75 year old male (lot #0016; Mayo Clinic, Rochester MN). The frozen tissue was immediately homogenized and lysed using a Brinkmann Homogenizer Polytron-PT 3000 (Brinkmann Instruments, Inc. Westbury NY) in guanidinium isothiocyanate solution. The lysate was extracted once with phenol chloroform at pH 8.0 and once with acid phenol at pH 4.0 per

Stratagene's RNA isolation protocol (Stratagene Inc, San Diego CA). The RNA was precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in DEPC-treated water and DNase treated for 25 min at 37¡C. The reaction was stopped with an equal volume of acid phenol, and the RNA was isolated using the Qiagen Oligotex kit (QIAGEN Inc, Chatsworth CA) and used to construct the cDNA library.

The RNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (catalog #18248-013; Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (catalog #275105, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5a[a] competent cells (Cat. #18258-012, Gibco/BRL).

**Brain**

The BRAINOT03 cDNA library was constructed from normal brain tissue removed from a 26 year old male (lot #0003; Mayo Clinic, Rochester MN). The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury NJ). The reagents and extraction procedures were used as supplied in the Stratagene RNA Isolation Kit (Cat. # 200345; Stratagene). The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted once with phenol chloroform pH 8.0, once with acid phenol pH 4.0, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in water and DNase treated for 15 min at 37°C. The RNA was isolated using the Qiagen Oligotex kit (QIAGEN Inc, Chatsworth CA) and used to construct the cDNA library.

The RNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Cat. #18248-013; Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5a™ competent cells (Cat. #18258-012, Gibco/BRL).

**II Isolation and Sequencing of cDNA Clones**

Plasmid DNA was released from the cells and purified using the Miniprep Kit (Catalogue # 77468; Advanced Genetic Technologies Corporation, Gaithersburg MD). This kit consists of a 96 well block with reagents for 960 purifications. The recommended protocol was employed except for the following changes: 1) the 96 wells were each filled with only 1 ml of sterile

Terrific Broth (Catalog # 22711, LIFE TECHNOLOGIES[a], Gaithersburg MD) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) the bacteria were cultured for 24 hours after the wells were inoculated and then lysed with 60 µl of lysis buffer; 3) a centrifugation step employing the Beckman GS-6R @2900 rpm for 5 min was performed before the contents of the block were added to the primary filter plate; and 4) the optional step of adding isopropanol to TRIS buffer was not routinely performed. After the last step in the protocol, samples were transferred to a Beckman 96-well block for storage.

The cDNAs were sequenced by the method of Sanger F and AR Coulson (1975; J Mol Biol 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno NV) in combination with four Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown MA) and Applied Biosystems 377 or 373 DNA Sequencing Systems (Perkin Elmer), and reading frame was determined.

**III Homology Searching of cDNA Clones and Their Deduced Proteins**

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (TRW Inc, Los Angeles CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT- 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul SF (1993) J Mol Evol 36:290-300; Altschul, SF et al (1990) J Mol Biol 215:403-10), was used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is

especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches which do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

**IV Northern Analysis**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labelled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al. supra).

Analogous computer techniques using BLAST (Altschul SF 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte, Palo Alto CA). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

and it takes into acccount both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

**V Extension of HCDC-Encoding Polynucleotides to Full Length or to Recover Regulatory Elements**

Full length HCDC-encoding nucleic acid sequence (SEQ ID NO:2) is used to design

oligonucleotide primers for extending a partial nucleotide sequence to full length or for obtaining 5' sequences from genomic libraries. One primer is synthesized to initiate extension in the antisense direction (XLR) and the other is synthesized to extend sequence in the sense direction (XLF). Primers allow the extension of the known HCDC-encoding sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest (US Patent Application 08/487,112, filed June 7, 1995, specifically incorporated by reference). The initial primers are designed from the cDNA using OLIGO® 4.06 Primer Analysis Software (National Biosciences), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

The original, selected cDNA libraries, or a human genomic library are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired, additional sets of primers are designed to further extend the known region.

By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR is performed using the Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the following parameters:

| | |
|---|---|
| Step 1 | 94° C for 1 min (initial denaturation) |
| Step 2 | 65° C for 1 min |
| Step 3 | 68° C for 6 min |
| Step 4 | 94° C for 15 sec |
| Step 5 | 65° C for 1 min |
| Step 6 | 68° C for 7 min |
| Step 7 | Repeat step 4-6 for 15 additional cycles |
| Step 8 | 94° C for 15 sec |
| Step 9 | 65° C for 1 min |
| Step 10 | 68° C for 7:15 min |
| Step 11 | Repeat step 8-10 for 12 cycles |
| Step 12 | 72° C for 8 min |
| Step 13 | 4° C (and holding) |

A 5-10 $\mu$l aliquot of the reaction mixture is analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were selected and cut out of the gel. Further purification involves using a commercial gel extraction method such as

QIAQuick™ (QIAGEN Inc). After recovery of the DNA, Klenow enzyme was used to trim single-stranded, nucleotide overhangs creating blunt ends which facilitate religation and cloning.

After ethanol precipitation, the products are redissolved in 13 $\mu$l of ligation buffer, 1$\mu$l T4-DNA ligase (15 units) and 1$\mu$l T4 polynucleotide kinase are added, and the mixture is incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 $\mu$l of appropriate media) are transformed with 3 $\mu$l of ligation mixture and cultured in 80 $\mu$l of SOC medium (Sambrook J et al, supra). After incubation for one hour at 37° C, the whole transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook J et al, supra) containing 2xCarb. The following day, several colonies are randomly picked from each plate and cultured in 150 $\mu$l of liquid LB/2xCarb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 $\mu$l of each overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 $\mu$l of each sample is transferred into a PCR array.

For PCR amplification, 18 $\mu$l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction are added to each well. Amplification is performed using the following conditions:

| | |
|---|---|
| Step 1 | 94° C for 60 sec |
| Step 2 | 94° C for 20 sec |
| Step 3 | 55° C for 30 sec |
| Step 4 | 72° C for 90 sec |
| Step 5 | Repeat steps 2-4 for an additional 29 cycles |
| Step 6 | 72° C for 180 sec |
| Step 7 | 4° C (and holding) |

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products are compared to the original partial cDNAs, and appropriate clones are selected, ligated into plasmid and sequenced.

**VI Labeling and Use of Hybridization Probes**

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 mCi of [$\gamma$-$^{32}$P] adenosine triphosphate (Amersham, Chicago IL) and T4 polynucleotide kinase (DuPont NEN®,

Boston MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 super fine resin column (Pharmacia). A portion containing $10^7$ counts per minute of each of the sense and antisense oligonucleotides is used in a typical membrane based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale CA) for several hours, hybridization patterns are compared visually.

**VII Antisense Molecules**

The HCDC-encoding sequence, or any part thereof, is used to inhibit in vivo or in vitro expression of naturally occurring HCDC. Although use of antisense oligonucleotides, comprising about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. An oligonucleotide based on the coding sequences of HCDC, as shown in Figures 1A, 1B, 1C, 1D, 2A, 2B, 2C and 2D is used to inhibit expression of naturally occurring HCDC. The complementary oligonucleotide is designed from the most unique 5' sequence as shown in Figures 1A, 1B, 1C, 1D, 2A, 2B, 2C and 2D and used either to inhibit transcription by preventing promoter binding to the upstream nontranslated sequence or translation of an HCDC-encoding transcript by preventing the ribosome from binding. Using an appropriate portion of the leader and 5' sequence of SEQ ID NO:2, an effective antisense oligonucleotide includes any 15-20 nucleotides spanning the region which translates into the signal or early coding sequence of the polypeptide as shown in Figures 1A, 1B, 1C, 1D, 2A, 2B, 2C and 2D.

**VIII Expression of HCDC**

Expression of the HCDC is accomplished by subcloning the cDNAs into appropriate vectors and transfecting the vectors into host cells. In this case, the cloning vector, pSport, previously used for the generation of the cDNA library is used to express HCDC in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met and the subsequent 7 residues of ß-galactosidase.

Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transfected bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first seven residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length HCDC-encoding sequence. The signal sequence directs the secretion of HCDC into the bacterial growth media which can be used directly in the following assay for activity.

**IX HCDC Activity**

Some mammalian homologs of yeast cdc genes can complement the respective yeast cdc mutants (Ninomiya-Tsu J et al (1991) Proc Natl Acad Sci 88: 9006-9010). HCDC complementation activity can be measured in yeast cells be methods described by Ninomiya-Tsu et al (supra). The HCDC gene is placed on an expression vector and transformed into either a Cdc36 or a Cdc37 temperature-sensitive mutant yeast strain. Growth of the yeast cells at the restrictive temperature indicates HCDC complementation activity.

HCDCA activity can also be assayed by a method described by Grammatikakis et al (supra). Extracts of bacterial cells expressing HCDCA are used to make western blots (Towbin H et al (1979) Proc Natl Acad Sci 76: 4350-4354). Western blots can be reacted with [$^3$H] hyaluronan as described by Banerjee SD et al (1991, Dev Biol 146: 186-197). Autoradiography reveals hyaluronan binding activity.

**X Production of HCDC Specific Antibodies**

HCDC substantially purified using PAGE electrophoresis (Sambrook, supra) is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence translated from HCDC is analyzed using DNAStar software (DNAStar Inc) to determine regions of high immunogenicity and a corresponding oligopolypeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Analysis to select appropriate epitopes, such as those near the C-terminus or in hydrophilic regions (shown in Figures 7 and 9) is described by Ausubel FM et al (supra).

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma) by reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel FM et al, supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity,

for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated, goat anti-rabbit IgG.

**XI Purification of Naturally Occurring HCDC Using Specific Antibodies**

Naturally occurring or recombinant HCDC is substantially purified by immunoaffinity chromatography using antibodies specific for HCDC. An immunoaffinity column is constructed by covalently coupling HCDC antibody to an activated chromatographic resin such as CnBr-activated Sepharose (Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HCDC is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HCDC (eg, high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HCDC binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope such as urea or thiocyanate ion), and HCDC is collected.

**XII Identification of Molecules Which Interact with HCDC**

HCDC, or biologically active fragments thereof, are labelled with $^{125}$I Bolton-Hunter reagent (Bolton, AE and Hunter, WM (1973) Biochem J 133: 529). Candidate molecules previously arrayed in the wells of a 96 well plate are incubated with the labelled HCDC, washed and any wells with labelled HCDC complex are assayed. Data obtained using different concentrations of HCDC are used to calculate values for the number, affinity, and association of HCDC with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

PF-0122 PCT

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF THE INVENTION: NOVEL HUMAN CELL DIVISION CYCLE PROTEINS

(iii) NUMBER OF SEQUENCES: 9

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
(B) STREET: 3174 Porter Drive
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: U.S.
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) PCT APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Filed Herewith

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/712,708
(B) FILING DATE: 12-SEP-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Billings, Lucy J.
(B) REGISTRATION NUMBER: 36,749
(C) REFERENCE/DOCKET NUMBER: PF-0122 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 650-855-0555
(B) TELEFAX: 650-845-4166

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 378 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY:
(B) CLONE: Consensus

PF-0122 PCT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Met Val Asp Tyr Ser Val Trp Asp His Ile Glu Val Ser Asp Asp Glu
1               5                   10                  15
Asp Glu Thr His Pro Asn Ile Asp Thr Ala Ser Leu Phe Arg Trp Arg
            20                  25                  30
His Gln Ala Arg Val Glu Arg Met Glu Gln Phe Gln Lys Glu Lys Glu
        35                  40                  45
Glu Leu Asp Arg Gly Cys Arg Glu Cys Lys Arg Lys Val Ala Glu Cys
    50                  55                  60
Gln Arg Lys Leu Lys Glu Leu Glu Val Ala Glu Gly Gly Lys Ala Glu
65                  70                  75                  80
Leu Glu Arg Leu Gln Ala Glu Ala Gln Gln Leu Arg Lys Glu Glu Arg
                85                  90                  95
Ser Trp Glu Gln Lys Leu Glu Glu Met Arg Lys Lys Glu Lys Ser Met
            100                 105                 110
Pro Trp Asn Val Asp Thr Leu Ser Lys Asp Gly Phe Ser Lys Ser Met
        115                 120                 125
Val Asn Thr Lys Pro Glu Lys Thr Glu Glu Asp Ser Glu Glu Val Arg
    130                 135                 140
Glu Gln Lys His Lys Thr Phe Val Glu Lys Tyr Glu Lys Gln Ile Lys
145                 150                 155                 160
His Phe Gly Met Leu Arg Arg Trp Asp Asp Ser Gln Lys Tyr Leu Ser
                165                 170                 175
Asp Asn Val His Leu Val Cys Glu Glu Thr Ala Asn Tyr Leu Val Ile
            180                 185                 190
Trp Cys Ile Asp Leu Glu Val Glu Glu Lys Cys Ala Leu Met Glu Gln
        195                 200                 205
Val Ala His Gln Thr Ile Val Met Gln Phe Ile Leu Glu Leu Ala Lys
    210                 215                 220
Ser Leu Lys Val Asp Pro Arg Ala Cys Phe Arg Gln Phe Phe Thr Lys
225                 230                 235                 240
Ile Lys Thr Ala Asp Arg Gln Tyr Met Glu Gly Phe Asn Asp Glu Leu
                245                 250                 255
Glu Ala Phe Lys Glu Arg Val Arg Gly Arg Ala Lys Leu Arg Ile Glu
            260                 265                 270
Lys Ala Met Lys Glu Tyr Glu Glu Glu Glu Arg Lys Lys Arg Leu Gly
        275                 280                 285
Pro Gly Gly Leu Asp Pro Val Glu Val Tyr Glu Ser Leu Pro Glu Glu
    290                 295                 300
Leu Gln Lys Cys Phe Asp Val Lys Asp Val Gln Met Leu Gln Asp Ala
305                 310                 315                 320
Ile Ser Lys Met Asp Pro Thr Asp Ala Lys Tyr His Met Gln Arg Cys
                325                 330                 335
Ile Asp Ser Gly Leu Trp Val Pro Asn Ser Lys Ala Ser Glu Ala Lys
            340                 345                 350
Glu Gly Glu Glu Ala Gly Pro Gly Asp Pro Leu Leu Glu Ala Val Pro
        355                 360                 365
Lys Thr Gly Asp Glu Lys Asp Val Ser Val
    370                 375
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

PF-0122 PCT

(vii) IMMEDIATE SOURCE:
(A) LIBRARY:
(B) CLONE: Consensus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
TCGTTTTATC GTCGCCCTCT CTCAAGCCGG AGCGGGCTGG CCCCCAAGGC AAATGGTGGA     60
CTACAGCGTG TGGGACCACA TTGAGGTGTC TGATGATGAA GACGAGACGC ACCCCAACAT    120
CGACACGGCC AGTCTCTTCC GCTGGCGGCA TCAGGCCCGG GTGGAACGCA TGGAGCAGTT    180
CCAGAAGGAG AAGGAGGAAC TGGACAGGGG CTGCCGCGAG TGCAAGCGCA AGGTGGCCGA    240
GTGCCAGAGG AAACTGAAGG AGCTGGAGGT GGCCGAGGGC GGCAAGGCAG AGCTGGAGCG    300
CCTGCAGGCC GAGGCACAGC AGCTGCGCAA GGAGGAGCGG AGCTGGGAGC AGAAGCTGGA    360
GGAGATGCGC AAGAAGGAGA AGAGCATGCC CTGGAACGTG GACACGCTCA GCAAAGACGG    420
CTTCAGCAAG AGCATGGTAA ATACCAAGCC CGAGAAGACG GAGGAGGACT CAGAGGAGGT    480
GAGGGAGCAG AAACACAAGA CCTTCGTGGA AAAATACGAG AAACAGATCA AGCACTTTGG    540
CATGCTTCGC CGCTGGGATG ACAGCCAAAA GTACCTGTCA GACAACGTCC ACCTGGTGTG    600
CGAGGAGACA GCCAATTACC TGGTCATTTG GTGCATTGAC CTAGAGGTGG AGGAGAAATG    660
TGCACTCATG GAGCAGGTGG CCCACCAGAC AATCGTCATG CAATTTATCC TGGAGCTGGC    720
CAAGAGCCTA AAGGTGGACC CCCGGGCCTG CTTCCGGCAG TTCTTCACTA AGATTAAGAC    780
AGCCGATCGC CAGTACATGG AGGGCTTCAA CGACGAGCTG GAAGCCTTCA AGGAGCGTGT    840
GCGGGGCCGT GCCAAGCTGC GCATCGAGAA GGCCATGAAG GAGTACGAGG AGGAGGAGCG    900
CAAGAAGCGG CTCGGCCCCG GCGGCCTGGA CCCCGTCGAG GTCTACGAGT CCCTCCCTGA    960
GGAACTCCAG AAGTGCTTCG ATGTGAAGGA CGTGCAGATG CTGCAGGACG CCATCAGCAA   1020
GATGGACCCC ACCGACGCAA AGTACCACAT GCAGCGCTGC ATTGACTCTG GCCTCTGGGT   1080
CCCCAACTCT AAGGCCAGCG AGGCCAAGGA GGGAGAGGAG GCAGGTCCTG GGGACCCATT   1140
ACTGGAAGCT GTTCCCAAGA CGGGCGATGA GAAGGATGTC AGTGTGTGAC CTGCCCCAGC   1200
TACCAMCGCC AGCTGCTTYC AGGGCCCTAT GTGCCCCTTT TCAGAAAACA GATAGATGCC   1260
ATCTCGCCCG CTCCTGACTT CCTCTACTTG CGCTGCTCGG CCCAACCTGG GGGGCCCGCC   1320
CAACCCTCCC TGGCCTCTCC ACTGTCTCCA CTCTCCAGCG CCCATTCAAG TCCCTGCTTT   1380
GAGTCAAGGG GCTTCACTGC CTGCAGCCCC CCATCAGCAT TATTCCAAAG GCCCGGGGGT   1440
CCGGGGAAGG GCAAAGGTCC CCAGGCTGGT CTCCCAGGTA GTTGGGGAGG GTCCCCANCC   1500
AAGGGGCCGG CTCCCGTCAC TGGGCCCTGT TTTCACTGTT CGTCTGCTGT CTGTGTCCTC   1560
TATTTGGCAA ACAGCAATGA TCTTCCAATA AAAGATTTCA GATGCCC                1607
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 280 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY:
(B) CLONE: Consensus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Val Thr Lys Pro Ala Asn Glu Gln Ser Gln Asp Phe Ser Ile His
 1               5                  10                  15
Asn Glu Asp Phe Pro Ala Leu Pro Gly Ser Ser Tyr Lys Asp Pro Thr
            20                  25                  30
Ser Ser Asn Asp Asp Ser Lys Ser Asn Leu Asn Thr Ser Gly Lys Thr
        35                  40                  45
Thr Ser Ser Thr Asp Gly Pro Lys Phe Pro Gly Asp Lys Ser Ser Thr
    50                  55                  60
Thr Gln Asn Asn Asn Gln Gln Lys Lys Gly Ile Gln Val Leu Pro Asp
65                  70                  75                  80
```

PF-0122 PCT

```
Gly Arg Val Thr Asn Ile Pro Gln Gly Met Val Thr Asp Gln Phe Gly
                85                  90                  95
Met Ile Gly Leu Leu Thr Phe Ile Arg Ala Ala Glu Thr Asp Pro Gly
            100                 105                 110
Met Val His Leu Ala Leu Gly Ser Asp Leu Thr Thr Leu Gly Leu Asn
        115                 120                 125
Leu Asn Ser Pro Glu Asn Leu Tyr Pro Lys Phe Ala Ser Pro Trp Ala
    130                 135                 140
Ser Ser Pro Cys Arg Pro Gln Asp Ile Asp Phe His Val Pro Ser Glu
145                 150                 155                 160
Tyr Leu Thr Asn Ile His Ile Arg Asp Lys Leu Ala Ala Ile Lys Leu
                165                 170                 175
Gly Arg Tyr Gly Glu Asp Leu Leu Phe Tyr Leu Tyr Tyr Met Asn Gly
            180                 185                 190
Gly Asp Val Leu Gln Leu Leu Ala Ala Val Glu Leu Phe Asn Arg Asp
        195                 200                 205
Trp Arg Tyr His Lys Glu Glu Arg Val Trp Ile Thr Arg Ala Pro Gly
    210                 215                 220
Met Glu Pro Thr Met Lys Thr Asn Thr Tyr Glu Arg Gly Thr Tyr Tyr
225                 230                 235                 240
Phe Phe Asp Cys Leu Xaa Trp Arg Lys Val Ala Lys Glu Phe His Leu
                245                 250                 255
Glu Tyr Asp Lys Leu Glu Glu Arg Pro His Leu Pro Ser Thr Phe Asn
            260                 265                 270
Tyr Asn Pro Ala Gln Gln Ala Phe
        275                 280
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1309 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY:
(B) CLONE: Consensus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
TNTCNTTTAN CACGGACGCG TGGGNGGCC CCCTGGGAAA AAATGTCACT TNNCACGCCT     60
CCATCTCCAA GCAGGGGTAT TTGGCCTCTG AATCCTAGGA ATATGATGAA CCACTCCCAG    120
GTTGGTCAGG GCNTTGGAAT TCCTAGCAGG ACAAATAGCA TGAGCAGTTC ANGGTTAGGT    180
AGCCCCAACA GAANCTCGCC AAGCATAATA TGTNTNCCNA AGCAGCAGCC TTCTCGACAG    240
CCTTTTACTG TGAACAGTAT GTCTGGATTT GGAATGAACA GGAATCAGGC ATTTGGAATG    300
AATAACTCCT TATCAAGTAA CATTTTTNTT NNANCANACG GAANTGAAAA TGTGACAGGA    360
TTGGACCTTT CAGATTTCCC ANCATTANCA GACCGAAACA GGAGGGAAGG AAGTGGTAAC    420
CCAACTCCAT TAATAAACCC CTTGGCTGGA ANAGCTCCTT ATNTTGGAAT GGTAACAAAA    480
CCAGCAAATG AACAATCCCA GGACTTCTCA ATACACAATG AAGATTTTCC AGCATTACCA    540
GGNTCCAGCT ATAAAGATCC AACATCAAGT AATGATGACA GTAAATCTAA TTTGAATACA    600
TCTGGCAAGA CAACTTCAAG TACAGATGGA CCCAAATTCC CTGGAGATAA AAGTTCAACA    660
ACACAAAATA ATAACCAGCA GAAAAAAGGG ATCCAGGTGT TACCTGATGG TCGGGTTACT    720
AACATTCCTC AAGGGATGGT GACGGACCAA TTTGGAATGA TTGGCCTGTT AACATTTATC    780
AGGGCAGCAG AGACAGACCC AGGAATGGTA CATCTTGCAT TAGGAAGTGA CTTAACAACA    840
TTAGGCCTCA ATCTGAACTC TCCTGAAAAT CTCTACCCCA AATTTGCGTC ACCCTGGGCA    900
TCTTCACCTT GTCGACCTCA AGACATAGAC TTCCATGTTC CATCTGAGTA CTTAACGAAC    960
ATTCACATTA GGGATAAGCT GGCTGCAATA AAACTTGGCC GATATGGTGA AGACCTTCTC   1020
```

PF-0122 PCT

```
TTCTATCTCT ATTACATGAA TGGAGGAGAC GTATTACAAC TTTTAGCTGC AGTGGAGCTT    1080
TTTAACCGTG ATTGGAGATA CCACAAAGAA GAACGAGTAT GGATTACCAG GGCACCAGGC    1140
ATGGAGCCAA CAATGAAAAC CAATACCTAT GAGAGGGGAA CATATTACTT CTTTGACTGT    1200
CTTAANTGGA GGAAAGTAGC TAAGGAGTTC CATCTGGAAT ATGACAAATT AGAAGAACGG    1260
CCTCACCTGC CATCCACCTT CAACTACAAC CCTGCTCAGC AAGCCTTCT                1309
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 246 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 755484

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Glu Glu Leu Arg Lys Lys Glu Lys Asn Met Pro Trp Asn Val Asp
 1               5                  10                  15
Thr Leu Ser Lys Asp Gly Phe Ser Lys Ser Val Phe Lys Leu Lys Ala
            20                  25                  30
Glu Glu Lys Glu Glu Thr Glu Glu Gln Lys Glu Gln Lys His Lys Thr
        35                  40                  45
Phe Val Glu Arg His Glu Lys Gln Ile Lys His Phe Gly Met Leu Arg
    50                  55                  60
Arg Trp Asp Asp Ser Gln Lys Tyr Leu Ser Asp Asn Pro His Leu Val
65                  70                  75                  80
Cys Glu Glu Thr Ala Asn Tyr Leu Val Ile Trp Cys Ile Asp Leu Glu
                85                  90                  95
Val Glu Glu Lys Gln Ala Leu Met Glu Gln Val Ala His Gln Thr Ile
            100                 105                 110
Val Met Gln Phe Ile Leu Glu Leu Ala Lys Ser Leu Lys Val Asp Pro
        115                 120                 125
Arg Ala Cys Phe Arg Gln Phe Phe Thr Lys Ile Lys Thr Ala Asp Gln
    130                 135                 140
Gln Tyr Met Glu Gly Phe Asn Asp Glu Leu Glu Ala Phe Lys Glu Arg
145                 150                 155                 160
Val Arg Gly Arg Ala Lys Ala Arg Ile Glu Arg Ala Met Arg Glu Tyr
                165                 170                 175
Glu Glu Glu Glu Arg Gln Lys Arg Leu Gly Pro Gly Gly Leu Asp Pro
            180                 185                 190
Val Asp Val Tyr Glu Ser Leu Pro Pro Glu Leu Gln Lys Cys Phe Asp
        195                 200                 205
Ala Lys Asp Val Gln Met Leu Gln Asp Thr Ile Ser Arg Met Asp Pro
    210                 215                 220
Thr Glu Ala Lys Tyr His Met Gln Arg Cys Ile Asp Ser Gly Leu Trp
225                 230                 235                 240
Val Pro Thr Gln His Gln
                245
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 379 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

PF-0122 PCT

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 1197180

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Val Asp Tyr Ser Val Trp Asp His Ile Glu Val Ser Asp Asp Glu
 1               5                  10                  15
Asp Glu Thr His Pro Asn Ile Asp Thr Ala Ser Leu Phe Arg Trp Arg
            20                  25                  30
His Gln Ala Arg Val Glu Arg Met Glu Gln Phe Gln Lys Glu Lys Glu
        35                  40                  45
Glu Leu Asp Arg Gly Cys Arg Glu Cys Lys Arg Lys Val Ala Glu Phe
    50                  55                  60
Gln Arg Lys Leu Lys Glu Leu Glu Val Ala Glu Gly Gly Gly Gln Val
65                  70                  75                  80
Glu Leu Glu Arg Leu Arg Ala Glu Ala Gln Gln Leu Arg Lys Glu Glu
                85                  90                  95
Arg Thr Gly Ser Arg Ser Trp Arg Thr Cys Gly Lys Lys Glu Lys Asn
            100                 105                 110
Met Pro Trp Asn Val Asp Thr Leu Ser Lys Asp Gly Phe Ser Lys Ser
        115                 120                 125
Met Val Asn Thr Lys Pro Glu Lys Ala Glu Glu Asp Ser Glu Glu Ala
    130                 135                 140
Arg Glu Gln Lys His Lys Thr Phe Val Glu Lys Tyr Glu Lys Gln Ile
145                 150                 155                 160
Lys His Phe Gly Met Leu His Arg Trp Asp Asp Ser Gln Lys Tyr Leu
                165                 170                 175
Ser Asp Asn Val His Leu Val Cys Glu Glu Thr Ala Asn Tyr Leu Val
            180                 185                 190
Ile Trp Cys Ile Asp Leu Glu Val Glu Glu Lys Cys Ala Leu Met Glu
        195                 200                 205
Gln Val Ala His Gln Thr Met Val Met Gln Phe Ile Leu Glu Leu Ala
    210                 215                 220
Lys Ser Leu Lys Val Asp Pro Arg Ala Cys Phe Arg Gln Phe Phe Thr
225                 230                 235                 240
Lys Ile Lys Thr Ala Asp Gln Gln Tyr Met Glu Gly Phe Lys Tyr Glu
                245                 250                 255
Leu Glu Ala Phe Lys Glu Arg Val Arg Gly Arg Ala Lys Leu Arg Ile
            260                 265                 270
Glu Lys Ala Met Lys Glu Tyr Glu Glu Glu Glu Arg Lys Lys Arg Leu
        275                 280                 285
Gly Pro Gly Gly Leu Asp Pro Val Glu Val Tyr Glu Ser Leu Pro Glu
    290                 295                 300
Glu Leu Gln Lys Cys Phe Asp Val Lys Asp Val Gln Met Leu Gln Asp
305                 310                 315                 320
Ala Ile Ser Lys Met Asp Pro Thr Asp Ala Lys Tyr His Met Gln Arg
                325                 330                 335
Cys Ile Asp Ser Gly Leu Trp Val Pro Asn Ser Lys Ser Gly Glu Ala
            340                 345                 350
Lys Glu Gly Glu Glu Ala Gly Pro Gly Asp Pro Leu Leu Glu Ala Val
        355                 360                 365
Pro Lys Ala Gly Phe Glu Lys Asp Ile Ser Ala
    370                 375
```

(2) INFORMATION FOR SEQ ID NO:7:

PF-0122 PCT

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 506 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 1077057

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Ala Ile Asp Tyr Ser Lys Trp Asp Lys Ile Glu Leu Ser Asp Asp
 1               5                  10                  15
Ser Asp Val Glu Val His Pro Asn Val Asp Lys Lys Ser Phe Ile Lys
            20                  25                  30
Trp Lys Gln Gln Ser Ile His Glu Gln Arg Phe Lys Arg Asn Gln Asp
        35                  40                  45
Ile Lys Asn Leu Glu Thr Gln Val Asp Met Tyr Ser His Leu Asn Lys
    50                  55                  60
Arg Val Asp Arg Ile Leu Ser Asn Leu Pro Glu Ser Ser Leu Thr Asp
65                  70                  75                  80
Leu Pro Ala Val Thr Lys Phe Leu Asn Ala Asn Phe Asp Lys Met Glu
                85                  90                  95
Lys Ser Lys Gly Glu Asn Val Asp Pro Glu Ile Ala Thr Tyr Asn Glu
            100                 105                 110
Met Val Glu Asp Leu Phe Glu Gln Leu Ala Lys Asp Leu Asp Lys Glu
        115                 120                 125
Gly Lys Asp Ser Lys Ser Pro Ser Leu Ile Arg Asp Ala Ile Leu Lys
    130                 135                 140
His Arg Ala Lys Ile Asp Ser Val Thr Val Glu Ala Lys Lys Lys Leu
145                 150                 155                 160
Asp Glu Leu Tyr Lys Glu Lys Asn Ala His Ile Ser Ser Glu Asp Ile
                165                 170                 175
His Thr Gly Phe Asp Ser Ser Phe Met Asn Lys Gln Lys Gly Gly Ala
            180                 185                 190
Lys Pro Leu Glu Ala Thr Pro Ser Glu Ala Leu Ser Ser Ala Ala Glu
        195                 200                 205
Ser Asn Ile Leu Asn Lys Leu Ala Lys Ser Ser Val Pro Gln Thr Phe
    210                 215                 220
Ile Asp Phe Lys Asp Asp Pro Met Lys Leu Ala Lys Glu Thr Glu Glu
225                 230                 235                 240
Phe Gly Lys Ile Ser Ile Asn Glu Tyr Ser Lys Ser Gln Lys Phe Leu
                245                 250                 255
Leu Glu His Leu Pro Ile Ile Ser Glu Gln Gln Lys Asp Ala Leu Met
            260                 265                 270
Met Lys Ala Phe Glu Tyr Gln Leu His Gly Asp Asp Lys Met Thr Leu
        275                 280                 285
Gln Val Ile His Gln Ser Glu Leu Met Ala Tyr Ile Lys Glu Ile Tyr
    290                 295                 300
Asp Met Lys Lys Ile Pro Tyr Leu Asn Pro Met Glu Leu Ser Asn Val
305                 310                 315                 320
Ile Asn Met Phe Phe Glu Lys Val Ile Phe Asn Lys Asp Lys Pro Met
                325                 330                 335
Gly Lys Glu Ser Phe Leu Arg Ser Val Gln Glu Lys Phe Leu His Ile
            340                 345                 350
Gln Lys Arg Ser Lys Ile Leu Gln Gln Glu Glu Met Asp Glu Ser Asn
        355                 360                 365
```

PF-0122 PCT

```
Ala Glu Gly Val Glu Thr Ile Gln Leu Lys Ser Leu Asp Asp Ser Thr
    370                 375                 380
Glu Leu Glu Val Asn Leu Pro Asp Phe Asn Ser Lys Asp Pro Glu Glu
385                 390                 395                     400
Met Lys Lys Val Lys Val Phe Lys Thr Leu Ile Pro Glu Lys Met Gln
                405                 410                 415
Glu Ala Ile Met Thr Lys Asn Leu Asp Asn Ile Asn Lys Val Phe Glu
            420                 425                 430
Asp Ile Pro Ile Glu Glu Ala Glu Lys Leu Leu Glu Val Phe Asn Asp
        435                 440                 445
Ile Asp Ile Ile Gly Ile Lys Ala Ile Leu Glu Asn Glu Lys Asp Phe
    450                 455                 460
Gln Ser Leu Lys Asp Gln Tyr Glu Gln Asp His Glu Asp Ala Thr Met
465                 470                 475                     480
Glu Asn Leu Ser Leu Asn Asp Arg Asp Gly Gly Gly Asp Asn His Glu
                485                 490                 495
Glu Val Lys His Thr Ala Asp Thr Val Asp
            500                 505
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 1053220

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Asn Ser Val Gly Gly Val Ala Thr Glu Arg Arg Leu Pro Gln Thr
 1               5                  10                  15
Gln Gln Phe Leu Ser His Ser Asn Phe His Ser Asn Ala Thr Ile Ile
            20                  25                  30
Asp Glu Ser Gln Phe Pro Ser Leu Gly Ala Lys Gly Thr Ser Ser Leu
        35                  40                  45
Gly Gly Gly Gly Phe Ser Pro Ile Pro Thr Thr Ser Gly Gly Val Leu
    50                  55                  60
Asn Val Ala Gln Ser Ser Pro Ser Arg Asp Leu Tyr Gly Ala Gln Arg
65                  70                  75                  80
Pro Asn Tyr Ala Asn Leu Met Arg Ser Asp Pro Ser Leu Thr Asn Pro
                85                  90                  95
Glu Phe Gln Ile Gln Asn Glu Asp Phe Pro Ala Leu Pro Gly Val Gly
            100                 105                 110
Ser Gly Gln Thr Gln Arg Ser Met Leu Gly Asp Gln Leu Ala Asn Met
        115                 120                 125
Leu Ala Asp Asp His Gln Val Asp Phe Ala Gly Pro Leu Gly Asp Cys
    130                 135                 140
Asp Pro Ser Arg Leu Ser Gly Ile Ser Arg Asn Ser Gln Glu Gly Pro
145                 150                 155                 160
Met His Gly Ile Ile Thr His Pro Asp Gly Glu Val Thr Asn Ile Pro
                165                 170                 175
Ala Ser Met Leu Asp Asp Gln Phe Gly Met Ala Gly Leu Val Thr Tyr
            180                 185                 190
```

```
Leu Arg Thr Val Asp Asn Pro Ser Ile Val Ser Leu Ala Leu Gly Tyr
        195                 200                 205
Asp Leu Thr Thr Leu Gly Leu Asn Leu Asn Leu Ser Glu Arg Lys Leu
    210                 215                 220
Tyr Met Asn Phe Gly Gly Pro Trp Ala Asp Ser Pro Ile Arg Ala His
225                 230                 235                 240
Glu Leu Asp Val Lys Val Pro Glu Glu Tyr Met Thr His Asn His Ile
                245                 250                 255
Arg Asp Lys Leu Pro Pro Leu Arg Leu Asn Lys Val Ser Glu Asp Val
            260                 265                 270
Leu Phe Tyr Leu Phe Tyr Asn Cys Pro Asn Glu Ile Tyr Gln Val Ala
        275                 280                 285
Ala Ala Cys Glu Leu Tyr Ala Arg Glu Trp Arg Phe His Lys Ser Glu
    290                 295                 300
Gln Val Trp Leu Thr Arg Ser Gln Tyr Gly Gly Val Lys Glu Gln Thr
305                 310                 315                 320
Gly Asn Tyr Glu Lys Gly His Tyr Asn Val Phe Asp Gln Met Gln Trp
                325                 330                 335
Arg Lys Ile Pro Lys Glu Leu Lys Leu Glu Tyr Lys Glu Leu Glu Asp
            340                 345                 350
Arg Pro Lys Met Pro Gln Ser Val Ser Gly Gln Pro Thr Pro Tyr Lys
        355                 360                 365
Tyr Phe Phe Gln Gly Pro Gln Phe Pro Ser Gly Pro Glu Thr Gly Leu
    370                 375                 380
Met Leu Gln Met His Asn Leu Thr Leu Gly Thr Gly Gly Gly Gly Gly
385                 390                 395                 400
Gly Gln Ile Thr Pro Pro Ala Pro Ala Gly Leu Asn Gly Val Met Gly
                405                 410                 415
Gly Gly Gly Val Gly Ala Ala Gly Ile Gly Gly Ile Asn Val Gln Pro
            420                 425                 430
Gly Ala Val Pro Ser Ala Ala Arg Ala Thr Pro Asn
        435                 440
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 115930

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Met Glu Lys Phe Gly Leu Lys Ala Leu Val Pro Leu Leu Lys Leu Glu
 1               5                  10                  15
Asp Lys Glu Leu Ser Ser Thr Tyr Asp His Ser Met Thr Leu Gly Ala
            20                  25                  30
Asp Leu Ser Ser Met Leu Tyr Ser Leu Gly Ile Pro Arg Asp Ser Gln
        35                  40                  45
Asp His Arg Val Leu Asp Thr Phe Gln Ser Pro Trp Ala Glu Thr Ser
    50                  55                  60
Arg Ser Glu Val Glu Pro Arg Phe Phe Thr Pro Glu Ser Phe Thr Asn
65                  70                  75                  80
```

PF-0122 PCT

```
Ile Pro Gly Val Leu Gln Ser Thr Val Thr Pro Pro Cys Phe Asn Ser
                85                  90                  95
Ile Gln Asn Asp Gln Gln Arg Val Ala Leu Phe Gln Asp Glu Thr Leu
            100                 105                 110
Phe Phe Leu Phe Tyr Lys His Pro Gly Thr Val Ile Gln Glu Leu Thr
        115                 120                 125
Tyr Leu Glu Leu Arg Lys Arg Asn Trp Arg Tyr His Lys Thr Leu Lys
    130                 135                 140
Ala Trp Leu Thr Lys Asp Pro Met Met Glu Pro Ile Val Ser Ala Asp
145                 150                 155                 160
Gly Leu Ser Glu Arg Gly Ser Tyr Val Phe Phe Asp Pro Gln Arg Trp
                165                 170                 175
Glu Lys Cys Gln Arg Asp Phe Leu Leu Phe Tyr Asn Ala Ile Met
            180                 185                 190
```

## CLAIMS

1. A substantially purified human cell division cycle protein comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof.

2. An isolated and purified polynucleotide sequence encoding the protein of claim 1.

3. An isolated and purified polynucleotide sequence of claim 2 consisting of SEQ ID NO:2 or variants thereof.

4. A polynucleotide sequence which is complementary to SEQ ID NO:2 or variants thereof.

5. A recombinant expression vector containing the polynucleotide sequence of claim 2.

6. A recombinant host cell containing the vector of claim 5.

7. A method for producing a polypeptide comprising a polypeptide of SEQ ID NO:1, the method comprising the steps of:

a) culturing the host cell of claim 6 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

8. A pharmaceutical composition comprising a substantially purified human cell division cycle protein having an amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

9. A purified antibody which binds specifically to the polypeptide of claim 1.

10. A purified antagonist which specifically regulates or modulates the activity of the polypeptide of claim 1.

11. A pharmaceutical composition comprising a substantially purified antagonist of the polypeptide of claim 10 in conjunction with a suitable pharmaceutical carrier.

12. A method for treating cancer comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 11.

13. A substantially purified human cell division cycle protein comprising the amino acid sequence of SEQ ID NO:3 or fragments thereof.

14. An isolated and purified polynucleotide sequence encoding the protein of claim 13.

15. An isolated and purified polynucleotide sequence of claim 14 consisting of SEQ ID NO:4 or variants thereof.

16. A polynucleotide sequence which is complementary to SEQ ID NO:4 or variants thereof.

17. A recombinant expression vector containing the polynucleotide sequence of claim 14.

18. A recombinant host cell containing the vector of claim 17.

19. A method for producing a polypeptide comprising a polypeptide of SEQ ID NO:3, the method comprising the steps of:

a) culturing the host cell of claim 18 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

20. A pharmaceutical composition comprising a substantially purified human cell division cycle protein having an amino acid sequence of SEQ ID NO:3 in conjunction with a suitable pharmaceutical carrier.

21. A purified antibody which binds specifically to the polypeptide of claim 13.

22. A purified antagonist which specifically regulates or modulates the activity of the polypeptide of claim 13.

23. A pharmaceutical composition comprising a substantially purified antagonist of the polypeptide of claim 22 in conjunction with a suitable pharmaceutical carrier.

24. A method for treating cancer comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 23.

```
                9           18          27          36          45          54
5' CGT TTT ATC GTC GCC CTC TCT CAA GCC GGA GCG GGC TGG CCC CCA AGG CAA ATG
                                                                      M

               63          72          81          90          99         108
   GTG GAC TAC AGC GTG TGG GAC CAC ATT GAG GTG TCT GAT GAT GAA GAC GAG ACG
   V   D   Y   S   V   W   D   H   I   E   V   S   D   D   E   D   E   T

              117         126         135         144         153         162
   CAC CCC AAC ATC GAC ACG GCC AGT CTC TTC CGC TGG CGG CAT CAG GCC CGG GTG
   H   P   N   I   D   T   A   S   L   F   R   W   R   H   Q   A   R   V

              171         180         189         198         207         216
   GAA CGC ATG GAG CAG TTC CAG AAG GAG AAG GAG GAA CTG GAC AGG GGC TGC CGC
   E   R   M   E   Q   F   Q   K   E   K   E   E   L   D   R   G   C   R

              225         234         243         252         261         270
   GAG TGC AAG CGC AAG GTG GCC GAG TGC CAG AGG AAA CTG AAG GAG CTG GAG GTG
   E   C   K   R   K   V   A   E   C   Q   R   K   L   K   E   L   E   V

              279         288         297         306         315         324
   GCC GAG GGC GGC AAG GCA GAG CTG GAG CGC CTG CAG GCC GAG GCA CAG CAG CTG
   A   E   G   G   K   A   E   L   E   R   L   Q   A   E   A   Q   Q   L

              333         342         351         360         369         378
   CGC AAG GAG GAG CGG AGC TGG GAG CAG AAG CTG GAG GAG ATG CGC AAG AAG GAG
   R   K   E   E   R   S   W   E   Q   K   L   E   E   M   R   K   K   E
```

FIGURE 1A

```
        387         396         405         414         423         432
AAG AGC ATG CCC TGG AAC GTG GAC ACG CTC AGC AAA GAC GGC TTC AGC AAG AGC
K   S   M   P   W   N   V   D   T   L   S   K   D   G   F   S   K   S

        441         450         459         468         477         486
ATG GTA AAT ACC AAG CCC GAG AAG ACG GAG GAG GAC TCA GAG GAG GTG AGG GAG
M   V   N   T   K   P   E   K   T   E   E   D   S   E   E   V   R   E

        495         504         513         522         531         540
CAG AAA CAC AAG ACC TTC GTG GAA AAA TAC GAG AAA CAG ATC AAG CAC TTT GGC
Q   K   H   K   T   F   V   E   K   Y   E   K   Q   I   K   H   F   G

        549         558         567         576         585         594
ATG CTT CGC CGC TGG GAT GAC AGC CAA AAG TAC CTG TCA GAC AAC GTC CAC CTG
M   L   R   R   W   D   D   S   Q   K   Y   L   S   D   N   V   H   L

        603         612         621         630         639         648
GTG TGC GAG GAG ACA GCC AAT TAC CTG GTC ATT TGG TGC ATT GAC CTA GAG GTG
V   C   E   E   T   A   N   Y   L   V   I   W   C   I   D   L   E   V

        657         666         675         684         693         702
GAG GAG AAA TGT GCA CTC ATG GAG CAG GTG GCC CAC CAG ACA ATC GTC ATG CAA
E   E   K   C   A   L   M   E   Q   V   A   H   Q   T   I   V   M   Q

        711         720         729         738         747         756
TTT ATC CTG GAG CTG GCC AAG AGC CTA AAG GTG GAC CCC CGG GCC TGC TTC CGG
F   I   L   E   L   A   K   S   L   K   V   D   P   R   A   C   F   R
```

FIGURE 1B

```
        765         774         783         792         801         810
CAG TTC TTC ACT AAG ATT AAG ACA GCC GAT CGC CAG TAC ATG GAG GGC TTC AAC
Q   F   F   T   K   I   K   T   A   D   R   Q   Y   M   E   G   F   N

        819         828         837         846         855         864
GAC GAG CTG GAA GCC TTC AAG GAG CGT GTG CGG GGC CGT GCC AAG CTG CGC ATC
D   E   L   E   A   F   K   E   R   V   R   G   R   A   K   L   R   I

        873         882         891         900         909         918
GAG AAG GCC ATG AAG GAG TAC GAG GAG GAG GAG CGC AAG AAG CGG CTC GGC CCC
E   K   A   M   K   E   Y   E   E   E   E   R   K   K   R   L   G   P

        927         936         945         954         963         972
GGC GGC CTG GAC CCC GTC GAG GTC TAC GAG TCC CTC CCT GAG GAA CTC CAG AAG
G   G   L   D   P   V   E   V   Y   E   S   L   P   E   E   L   Q   K

        981         990         999        1008        1017        1026
TGC TTC GAT GTG AAG GAC GTG CAG ATG CTG CAG GAC GCC ATC AGC AAG ATG GAC
C   F   D   V   K   D   V   Q   M   L   Q   D   A   I   S   K   M   D

       1035        1044        1053        1062        1071        1080
CCC ACC GAC GCA AAG TAC CAC ATG CAG CGC TGC ATT GAC TCT GGC CTC TGG GTC
P   T   D   A   K   Y   H   M   Q   R   C   I   D   S   G   L   W   V

       1089        1098        1107        1116        1125        1134
CCC AAC TCT AAG GCC AGC GAG GCC AAG GAG GGA GAG GAG GCA GGT CCT GGG GAC
P   N   S   K   A   S   E   A   K   E   G   E   E   A   G   P   G   D
```

FIGURE 1C

```
       1143        1152        1161        1170        1179        1188
CCA TTA CTG GAA GCT GTT CCC AAG ACG GGC GAT GAG AAG GAT GTC AGT GTG TGA
P   L   L   E   A   V   P   K   T   G   D   E   K   D   V   S   V   *

       1197        1206        1215        1224        1233        1242
CCT GCC CCA GCT ACC AMC GCC AGC TGC TTY CAG GGC CCT ATG TGC CCC TTT TCA

       1251        1260        1269        1278        1287        1296
GAA AAC AGA TAG ATG CCA TCT CGC CCG CTC CTG ACT TCC TCT ACT TGC GCT GCT

       1305        1314        1323        1332        1341        1350
CGG CCC AAC CTG GGG GGC CCG CCC AAC CCT CCC TGG CCT CTC CAC TGT CTC CAC

       1359        1368        1377        1386        1395        1404
TCT CCA GCG CCC ATT CAA GTC CCT GCT TTG AGT CAA GGG GCT TCA CTG CCT GCA

       1413        1422        1431        1440        1449        1458
GCC CCC CAT CAG CAT TAT TCC AAA GGC CCG GGG GTC CGG GGA AGG GCA AAG GTC

       1467        1476        1485        1494        1503        1512
CCC AGG CTG GTC TCC CAG GTA GTT GGG GAG GGT CCC CAN CCA AGG GGC CGG CTC

       1521        1530        1539        1548        1557        1566
CCG TCA CTG GGC CCT GTT TTC ACT GTT CGT CTG CTG TCT GTG TCC TCT ATT TGG

       1575        1584        1593        1602
CAA ACA GCA ATG ATC TTC CAA TAA AAG ATT TCA GAT GCC C 3'
```

FIGURE 1D

```
                9           18          27          36          45          54
5' TNT CNT TTA NCA CGG ACG CGT GGG NGG GCC CCC TGG GAA AAA ATG TCA CTT NNC

               63          72          81          90          99         108
   ACG CCT CCA TCT CCA AGC AGG GGT ATT TGG CCT CTG AAT CCT AGG AAT ATG ATG

              117         126         135         144         153         162
   AAC CAC TCC CAG GTT GGT CAG GGC NTT GGA ATT CCT AGC AGG ACA AAT AGC ATG

              171         180         189         198         207         216
   AGC AGT TCA NGG TTA GGT AGC CCC AAC AGA ANC TCG CCA AGC ATA ATA TGT NTN

              225         234         243         252         261         270
   CCN AAG CAG CAG CCT TCT CGA CAG CCT TTT ACT GTG AAC AGT ATG TCT GGA TTT

              279         288         297         306         315         324
   GGA ATG AAC AGG AAT CAG GCA TTT GGA ATG AAT AAC TCC TTA TCA AGT AAC ATT

              333         342         351         360         369         378
   TTT NTT NNA NCA NAC GGA ANT GAA AAT GTG ACA GGA TTG GAC CTT TCA GAT TTC

              387         396         405         414         423         432
   CCA NCA TTA NCA GAC CGA AAC AGG AGG GAA GGA AGT GGT AAC CCA ACT CCA TTA

              441         450         459         468         477         486
   ATA AAC CCC TTG GCT GGA ANA GCT CCT TAT NTT GGA ATG GTA ACA AAA CCA GCA
                                                    M   V   T   K   P   A
```

FIGURE 2A

```
        495         504         513         522         531         540
AAT GAA CAA TCC CAG GAC TTC TCA ATA CAC AAT GAA GAT TTT CCA GCA TTA CCA
N   E   Q   S   Q   D   F   S   I   H   N   E   D   F   P   A   L   P

        549         558         567         576         585         594
GGN TCC AGC TAT AAA GAT CCA ACA TCA AGT AAT GAT GAC AGT AAA TCT AAT TTG
G   S   S   Y   K   D   P   T   S   S   N   D   D   S   K   S   N   L

        603         612         621         630         639         648
AAT ACA TCT GGC AAG ACA ACT TCA AGT ACA GAT GGA CCC AAA TTC CCT GGA GAT
N   T   S   G   K   T   T   S   S   T   D   G   P   K   F   P   G   D

        657         666         675         684         693         702
AAA AGT TCA ACA ACA CAA AAT AAT AAC CAG CAG AAA AAA GGG ATC CAG GTG TTA
K   S   S   T   T   Q   N   N   N   Q   Q   K   K   G   I   Q   V   L

        711         720         729         738         747         756
CCT GAT GGT CGG GTT ACT AAC ATT CCT CAA GGG ATG GTG ACG GAC CAA TTT GGA
P   D   G   R   V   T   N   I   P   Q   G   M   V   T   D   Q   F   G

        765         774         783         792         801         810
ATG ATT GGC CTG TTA ACA TTT ATC AGG GCA GCA GAG ACA GAC CCA GGA ATG GTA
M   I   G   L   L   T   F   I   R   A   A   E   T   D   P   G   M   V

        819         828         837         846         855         864
CAT CTT GCA TTA GGA AGT GAC TTA ACA ACA TTA GGC CTC AAT CTG AAC TCT CCT
H   L   A   L   G   S   D   L   T   T   L   G   L   N   L   N   S   P
```

FIGURE 2B

```
        873         882         891         900         909         918
GAA AAT CTC TAC CCC AAA TTT GCG TCA CCC TGG GCA TCT TCA CCT TGT CGA CCT
E   N   L   Y   P   K   F   A   S   P   W   A   S   S   P   C   R   P

        927         936         945         954         963         972
CAA GAC ATA GAC TTC CAT GTT CCA TCT GAG TAC TTA ACG AAC ATT CAC ATT AGG
Q   D   I   D   F   H   V   P   S   E   Y   L   T   N   I   H   I   R

        981         990         999        1008        1017        1026
GAT AAG CTG GCT GCA ATA AAA CTT GGC CGA TAT GGT GAA GAC CTT CTC TTC TAT
D   K   L   A   A   I   K   L   G   R   Y   G   E   D   L   L   F   Y

       1035        1044        1053        1062        1071        1080
CTC TAT TAC ATG AAT GGA GGA GAC GTA TTA CAA CTT TTA GCT GCA GTG GAG CTT
L   Y   Y   M   N   G   G   D   V   L   Q   L   L   A   A   V   E   L

       1089        1098        1107        1116        1125        1134
TTT AAC CGT GAT TGG AGA TAC CAC AAA GAA GAA CGA GTA TGG ATT ACC AGG GCA
F   N   R   D   W   R   Y   H   K   E   E   R   V   W   I   T   R   A

       1143        1152        1161        1170        1179        1188
CCA GGC ATG GAG CCA ACA ATG AAA ACC AAT ACC TAT GAG AGG GGA ACA TAT TAC
P   G   M   E   P   T   M   K   T   N   T   Y   E   R   G   T   Y   Y

       1197        1206        1215        1224        1233        1242
TTC TTT GAC TGT CTT AAN TGG AGG AAA GTA GCT AAG GAG TTC CAT CTG GAA TAT
F   F   D   C   L   X   W   R   K   V   A   K   E   F   H   L   E   Y
```

FIGURE 2C

```
 1251        1260        1269        1278        1287        1296
GAC AAA TTA GAA GAA CGG CCT CAC CTG CCA TCC ACC TTC AAC TAC AAC CCT GCT
D   K   L   E   E   R   P   H   L   P   S   T   F   N   Y   N   P   A

     1305
CAG CAA GCC TTC T 3'
Q   Q   A   F
```

FIGURE 2D

| Library | Lib Description | Abun | Pct Abun |
|---|---|---|---|
| TESTNOT04 | testis, 37 M | 2 | 0.2146 |
| SCORNON01 | spinal cord, 71 M, NORM | 1 | 0.1379 |
| LIVRTUT01 | liver tumor, metastasis, 51 F | 5 | 0.1294 |
| KERANOT02 | keratinocytes, primary cell line, 30 F | 3 | 0.1187 |
| TLYMNOT01 | lymphocytes (non-adher PBMNC), 24 M | 1 | 0.1059 |
| COLNNOT07 | colon, 60 M | 1 | 0.0920 |
| THP1PLB01 | THP-1 promonocyte cell line, treated PMA, LPS | 2 | 0.0903 |
| LUNGNOT04 | lung, 2 M | 4 | 0.0732 |
| COLNNOT09 | colon, 60 M | 1 | 0.0712 |
| PROSNOT07 | prostate, 69 M, match to PROSTUT05 | 2 | 0.0695 |
| SPLNNOT04 | spleen, 2 M | 5 | 0.0639 |
| BRAINOT14 | brain, 40 F, match to BRAITUT12 | 2 | 0.0629 |
| BEPINOT01 | bronchial epithelium, primary cell line, 54 M | 2 | 0.0609 |
| BRSTTUT03 | breast tumor, 58 F, match to BRSTNOT05 | 4 | 0.0593 |
| BRAITUT08 | brain tumor, astrocytoma, 47 M | 4 | 0.0585 |
| CRBLNOT01 | brain, cerebellum, 69 M | 3 | 0.0585 |
| PLACNOM01 | placenta, fetal M, WM | 1 | 0.0580 |
| TESTNOT03 | testis, 37 M | 1 | 0.0558 |
| BLADNOT04 | bladder, 28 M | 2 | 0.0555 |
| LUNGNOT10 | lung, fetal M | 2 | 0.0521 |
| PROSNOT14 | prostate, 60 M, match to PROSTUT08 | 2 | 0.0512 |
| BMARNOT03 | bone marrow, 16 M | 2 | 0.0484 |
| PROSNOT15 | prostate, 66 M, match to PROSTUT10 | 2 | 0.0483 |

FIGURE 3A

| | | | |
|---|---|---|---|
| KIDNNOT02 | kidney, 64 F | 1 | 0.0482 |
| PGANNOT01 | paraganglia, 46 M | 3 | 0.0479 |
| TMLR3DT01 | lymphocytes (non-adher PBMNC), M, 96-hr MLR | 2 | 0.0457 |
| BRAINOM02 | brain, 55 M, NORM, WM | 1 | 0.0454 |
| BRSTNOT03 | breast, 54 F, match to BRSTTUT02 | 3 | 0.0440 |
| PROSNOT02 | prostate, 50 M, match to PROSTUT01 | 1 | 0.0434 |
| PANCNOT01 | pancreas, 29 M | 2 | 0.0427 |
| THYRNOT03 | thyroid tumor, adenoma, 28 F | 3 | 0.0415 |
| THP1PLB02 | THP-1 promonocyte cell line, treated PMA, LPS | 1 | 0.0407 |
| UTRSNOT01 | uterus, 59 F | 1 | 0.0393 |
| THYMNOT02 | thymus, 3 M | 2 | 0.0386 |
| MUSCNOT02 | muscle, psoas, 12 M | 1 | 0.0382 |
| SINTTUT01 | small intestine tumor, 42 M | 1 | 0.0382 |
| TLYMNOR01 | lymphocytes (non-adher PBMNC), 24 M, RP | 1 | 0.0372 |
| STOMTUT01 | stomach tumor, 52 M, match to STOMNOT02 | 1 | 0.0367 |
| HUVESTB01 | HUVEC endothelial cell line, shear stress | 1 | 0.0359 |
| SYNOOAT01 | synovium, knee, osteoarthritis, 82 F | 2 | 0.0359 |
| BRAINOT04 | brain, choroid plexus, hemorrhage, 44 M | 1 | 0.0356 |
| PROSTUT03 | prostate tumor, 67 M, match to PROSNOT05 | 1 | 0.0352 |
| LVENNOT03 | heart, left ventricle, 31 M | 1 | 0.0337 |
| MMLR3DT01 | macrophages (adher PBMNC), M/F, 72-hr MLR | 1 | 0.0331 |
| THP1NOB01 | THP-1 promonocyte cell line, control | 1 | 0.0327 |
| OVARTUT01 | ovarian tumor, 43 F, match to OVARNOT03 | 1 | 0.0323 |
| BRAINOT11 | brain, right temporal, epilepsy, 5 M | 1 | 0.0322 |
| OVARNOT02 | ovary, 59 F | 1 | 0.0315 |

FIGURE 3B

| | | | |
|---|---|---|---|
| BRAITUT03 | brain tumor, astrocytoma, 17 F | 2 | 0.0307 |
| BRAINOT12 | brain, right frontal, epilepsy, 5 M | 1 | 0.0303 |
| PROSTUT05 | prostate tumor, 69 M, match to PROSNOT07 | 1 | 0.0303 |
| STOMNOT01 | stomach, 55 M | 1 | 0.0301 |
| COLNNOT16 | colon, 62 M, match to COLNTUT03 | 1 | 0.0295 |
| TONGTUT01 | tongue tumor, carcinoma, 36 M | 1 | 0.0295 |
| COLNNOT19 | large intestine, cecum, 18 F | 1 | 0.0293 |
| DUODNOT01 | small intestine, duodenum, 41 F | 1 | 0.0287 |
| PANCNOT07 | pancreas, fetal M | 1 | 0.0287 |
| UTRSNOT06 | uterus, myometrium, 50 F | 1 | 0.0282 |
| LUNGNOT12 | lung, 78 M | 1 | 0.0278 |
| LUNGNOT15 | lung, 69 M, match to LUNGTUT03 | 1 | 0.0276 |
| BEPINON01 | bronchial epithelium, primary cell line,54 M,NORM | 1 | 0.0274 |
| COLNTUT03 | colon tumor, 62 M, match to COLNNOT16 | 1 | 0.0272 |
| BRAINOM03 | brain, 55 M, NORM, WM | 1 | 0.0270 |
| BMARNOT02 | bone marrow, 16-70 M/F | 1 | 0.0269 |
| BRAITUT01 | brain tumor, oligoastrocytoma, 50 F | 2 | 0.0269 |
| KIDNNOT09 | kidney, fetal M | 1 | 0.0267 |
| PENITUT01 | penis tumor, carcinoma, 64 M | 1 | 0.0267 |
| PROSTUT08 | prostate tumor, 60 M, match to PROSNOT14 | 1 | 0.0266 |
| COLNNOT23 | colon, 16 M | 1 | 0.0264 |
| PROSTUT09 | prostate tumor, 66 M | 1 | 0.0264 |
| URETTUT01 | ureter tumor, 69 M | 1 | 0.0262 |
| LUNGNOT14 | lung, 47 M | 1 | 0.0259 |
| STOMFET01 | stomach, fetal F | 1 | 0.0255 |
| PROSNOT16 | prostate, 68 M | 1 | 0.0250 |
| MENITUT03 | brain tumor, benign meningioma, 35 F | 1 | 0.0249 |

FIGURE 3C

| | | | |
|---|---|---|---|
| COLNTUT02 | colon tumor, 75 M, match to COLNNOT01 | 1 | 0.0220 |
| SCORNOT01 | spinal cord, 71 M | 1 | 0.0201 |
| LUNGNOT03 | lung, 79 M, match to LUNGTUT02 | 1 | 0.0200 |
| HNT2AGT01 | hNT-2 cell line, post-mitotic neurons | 1 | 0.0190 |
| LUNGTUT02 | lung tumor, metastasis, 79 M, match to LUNGNOT03 | 1 | 0.0189 |
| BRAINOT03 | brain, 26 M | 1 | 0.0185 |
| HNT2RAT01 | hNT-2 cell line, teratocarcinoma, treated RA | 1 | 0.0185 |
| BRAITUT02 | brain tumor, metastasis, 58 M | 1 | 0.0169 |
| PANCNOT04 | pancreas, 5 M | 1 | 0.0169 |
| SINTBST01 | small intestine, ileum, Crohn's, 18 F | 1 | 0.0168 |
| BRSTNOT02 | breast, 55 F, match to BRSTTUT01 | 1 | 0.0158 |
| NGANNOT01 | ganglioneuroma, 9 M | 1 | 0.0155 |
| BRSTNOT05 | breast, 58 F, match to BRSTTUT03 | 1 | 0.0154 |
| CORPNOT02 | brain, corpus callosum, Alzheimer's, 74 M | 1 | 0.0153 |
| BRSTTUT01 | breast tumor, 55 F, match to BRSTNOT02 | 1 | 0.0151 |
| LUNGAST01 | lung, asthma, 17 M | 1 | 0.0150 |
| COLNFET02 | colon, fetal F | 1 | 0.0143 |
| BLADTUT04 | bladder tumor, 60 M, match to BLADNOT05 | 1 | 0.0127 |
| UCMCL5T01 | mononuclear cells, treated IL-5 | 1 | 0.0125 |
| EOSIHET02 | eosinophils, hypereosinophilia, 48 M | 1 | 0.0104 |
| BRAINOM01 | brain, infant F, NORM, WM | 1 | 0.0045 |
| LIVSFEM02 | liver/spleen, fetal M, NORM, WM | 1 | 0.0027 |

FIGURE 3D

| Library | Lib Description | Abun | Pct Abun |
|---|---|---|---|
| FIBRNGT01 | GD23A fibroblasts, radiation 5 min | 1 | 0.1664 |
| PITUNOR01 | pituitary, 16-70 M/F, RP | 1 | 0.1233 |
| MYOMNOT01 | uterus, myometrium, 43 F | 1 | 0.0409 |
| STOMTUT01 | stomach tumor, 52 M, match to STOMNOT02 | 1 | 0.0367 |
| BRAITUT02 | brain tumor, metastasis, 58 M | 2 | 0.0338 |
| STOMNOT02 | stomach, 52 M, match to STOMTUT01 | 1 | 0.0308 |
| LUNGNOT09 | lung, fetal M | 1 | 0.0286 |
| PTHYTUM01 | parathyroid tumor, adenoma, M/F, NORM, WM | 1 | 0.0278 |
| LNODNOT03 | lymph node, 67 M | 1 | 0.0265 |
| BRAITUT13 | brain tumor, meningioma, 68 M | 1 | 0.0262 |
| DUODNOT02 | small intestine, duodenum, 8 F | 1 | 0.0262 |
| BRAINOT03 | brain, 26 M | 1 | 0.0185 |
| HNT2RAT01 | hNT-2 cell line, teratocarcinoma, treated RA | 1 | 0.0185 |
| LUNGNOT04 | lung, 2 M | 1 | 0.0183 |
| UTRSNOT02 | uterus, 34 F | 1 | 0.0166 |
| NGANNOT01 | ganglioneuroma, 9 M | 1 | 0.0155 |
| BRAINOM01 | brain, infant F, NORM, WM | 3 | 0.0134 |
| UCMCL5T01 | mononuclear cells, treated IL-5 | 1 | 0.0125 |

FIGURE 4

```
1   M - V D Y S V W D H I E V S D D E D - E T H P N I D T A S L F R W R H Q A R V E   SEQ ID NO-1
1   M - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   SEQ ID NO-5
1   M - V D Y S V W D H I E V S D D E D - E T H P N I D T A S L F R W R H Q A R V E   SEQ ID NO-6
1   M A I D Y S K W D K I E L S D D S D V E V H P N V D K K S F I K W K Q Q S I H E   SEQ ID NO-7

39  R M E Q F Q K E K E E L D R G C R E C K R K V A E C Q R K L K E L - - - - - - -   SEQ ID NO-1
2   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   SEQ ID NO-5
39  R M E Q F Q K E K E E L D R G C R E C K R K V A E F Q R K L K E L - - - - - - -   SEQ ID NO-6
41  Q R F K R N Q D I K N L E T Q V D M Y S H L N K R V D R I L S N L P E S S L T D   SEQ ID NO-7

72  - - - - - - - - - - - - - - - - - - - - - - - - - E V A E G G - K A E - - L E R   SEQ ID NO-1
2   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   SEQ ID NO-5
72  - - - - - - - - - - - - - - - - - - - - - - - - - E V A E G G G Q V E - - L E R   SEQ ID NO-6
81  L P A V T K F L N A N F D K M E K S K G E N V D P E I A T Y N E M V E D L F E Q   SEQ ID NO-7

84  L Q A E A Q Q L R K E E R S W E Q K L E E M R K K E K S M P W - - - - - - - N V   SEQ ID NO-1
2   - - - - - E E L R K - - - - - - - - - - - - - - K E K N M P W - - - - - - - N V   SEQ ID NO-5
85  L R A E A Q Q L R K E E R T G S R S W R T C G K K E K N M P W - - - - - - - N V   SEQ ID NO-6
121 L A K D L D K E G K D S K S P S L I R D A I L K H R A K I D S V T V E A K K K L   SEQ ID NO-7

117 D T L S K D - - - - - - - - - - - - G F - - - - - - - - - - - S K S M V N T K P   SEQ ID NO-1
16  D T L S K D - - - - - - - - - - - - G F - - - - - - - - - - - S K S V F K L K A   SEQ ID NO-5
118 D T L S K D - - - - - - - - - - - - G F - - - - - - - - - - - S K S M V N T K P   SEQ ID NO-6
161 D E L Y K E K N A H I S S E D I H T G F D S S F M N K Q K G G A K P L E A T P S   SEQ ID NO-7
```

FIGURE 5A

```
134 E K T E E D S E E - - - - - - V R E Q K H K T F V E - - - - - - K Y E K Q I K H SEQ ID NO-1
33  E E K E E - T E E - - - - - - Q K E Q K H K T F V E - - - - - - R H E K Q I K H SEQ ID NO-5
135 E K A E E D S E E - - - - - - A R E Q K H K T F V E - - - - - - K Y E K Q I K H SEQ ID NO-6
201 E A L S S A A E S N I L N K L A K S S V P Q T F I D F K D D P M K L A K E T E E SEQ ID NO-7

162 F G M L - - R R W D D S Q K Y L S D N V H L V C E E T A N Y L V I W C I D L E V SEQ ID NO-1
60  F G M L - - R R W D D S Q K Y L S D N P H L V C E E T A N Y L V I W C I D L E V SEQ ID NO-5
163 F G M L - - H R W D D S Q K Y L S D N V H L V C E E T A N Y L V I W C I D L E V SEQ ID NO-6
241 F G K I S I N E Y S K S Q K F L L E H L P I I S E Q Q K D A L M M K A F E Y Q L SEQ ID NO-7

200 E E K C A L M E Q V A H Q T I V M Q F I L E L A K S L K V - - - - - - D P R A C SEQ ID NO-1
98  E E K Q A L M E Q V A H Q T I V M Q F I L E L A K S L K V - - - - - - D P R A C SEQ ID NO-5
201 E E K C A L M E Q V A H Q T M V M Q F I L E L A K S L K V - - - - - - D P R A C SEQ ID NO-6
281 H G D D K M T L Q V I H Q S E L M A Y I K E I Y D M K K I P Y L N P M E L S N V SEQ ID NO-7

234 F R Q F F T K I K T A D R Q Y M - - E G F N D E L E A F K E R V R G R A K L R I SEQ ID NO-1
132 F R Q F F T K I K T A D Q Q Y M - - E G F N D E L E A F K E R V R G R A K A R I SEQ ID NO-5
235 F R Q F F T K I K T A D Q Q Y M - - E G F K Y E L E A F K E R V R G R A K L R I SEQ ID NO-6
321 I N M F F E K V I F N K D K P M G K E S F L R S V Q E K F L H I Q K R S K I L Q SEQ ID NO-7

272 E K A M K E Y E E E E R K - - - - K R L G - - - - - - - - - - - - - - - - P G G SEQ ID NO-1
170 E R A M R E Y E E E E R Q - - - - K R L G - - - - - - - - - - - - - - - - P G G SEQ ID NO-5
273 E K A M K E Y E E E E R K - - - - K R L G - - - - - - - - - - - - - - - - P G G SEQ ID NO-6
361 Q E E M D E S N A E G V E T I Q L K S L D D S T E L E V N L P D F N S K D P E E SEQ ID NO-7
```

FIGURE 5B

```
292 L D P V E V Y E S L - P E E L Q K C F D V K D V Q M L Q D A I S K M D P T D A K  SEQ ID NO-1
190 L D P V D V Y E S L - P P E L Q K C F D A K D V Q M L Q D T I S R M D P T E A K  SEQ ID NO-5
293 L D P V E V Y E S L - P E E L Q K C F D V K D V Q M L Q D A I S K M D P T D A K  SEQ ID NO-6
401 M K K V K V F K T L I P E K M Q E A I M T K N L D N I N K V F E D I P I E E A E  SEQ ID NO-7

331 - - - - - Y H M Q R C I D S G L W V P N S K A S E A - K E G E E A G P G D P L L  SEQ ID NO-1
229 - - - - - Y H M Q R C I D S G L W V P T Q H Q                                      SEQ ID NO-5
332 - - - - - Y H M Q R C I D S G L W V P N S K S G E A - K E G E E A G P G D P L L  SEQ ID NO-6
441 K L L E V F N D I D I I G I K A I L E N E K D F Q S L K D Q Y E Q D H E D A T M  SEQ ID NO-7

365 E A V P K T G D E - - - - - - - - - - - - - K D V S V                            SEQ ID NO-1
246                                                                                    SEQ ID NO-5
366 E A V P K A G F E - - - - - - - - - - - - - K D I S A                            SEQ ID NO-6
481 E N L S L N D R D G G G D N H E E V K H T A D T V D                              SEQ ID NO-7
```

FIGURE 5C

```
1    M - - V T K P A N E - - - - Q S Q D F - - - - - - - - - - - S I H N E D F P A L   SEQ ID NO-3
1    M N S V G G V A T E R R L P Q T Q Q F L S H S N F H S N A T I I D E S Q F P S L   SEQ ID NO-8
1    M E K F G - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - L K A L   SEQ ID NO-9

24   P - - - - - - - - G S S Y K D P T S S N D D - - - - - - - - - - - - - - - - - -   SEQ ID NO-3
41   G A K G T S S L G G G G F S P I P T T S G G V L N V A Q S S P S R D L Y G A Q R   SEQ ID NO-8
10   V P - - - - - - - - - - - - - - - - - - - - L L K L E D - - - - K E L - - - - S   SEQ ID NO-9

38   - - - - - - - - S K S N L N T S G K T T S S T D G P K F P G D K S S T T Q - - -   SEQ ID NO-3
81   P N Y A N L M R S D P S L T N P E F Q I Q N E D F P A L P G V G S G Q T Q R S M   SEQ ID NO-8
22   S T Y D H S M T - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   SEQ ID NO-9

67   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - N N N Q Q   SEQ ID NO-3
121  L G D Q L A N M L A D D H Q V D F A G P L G D C D P S R L S G I S R N S Q E G P   SEQ ID NO-8
30   L G A D L S S M L - - - - - - - Y S - - - - - - - - - - - L G I P R D S Q D H R   SEQ ID NO-9

72   K K G I Q V L P D G R V T N I P Q G M V T D Q F G M I G L L T F I R A - A E T D   SEQ ID NO-3
161  M H G I I T H P D G E V T N I P A S M L D D Q F G M A G L V T Y L R T - V D - N   SEQ ID NO-8
52   V - - - - - - - - - - - - - - - - - - - - - - - - - - - L D T F Q S P W A E T S   SEQ ID NO-9

111  P G M V H L A L G S D L T T L G L N L N S P E - N L Y P K F A S P W A S S P C R   SEQ ID NO-3
199  P S I V S L A L G Y D L T T L G L N L N L S E R K L Y M N F G G P W A D S P I R   SEQ ID NO-8
65   R S E V - - - - - - - - - - - - - - - - - - E P R F F T - - P E S F T N I P - -   SEQ ID NO-9
```

FIGURE 6A

```
150 PQDIDFHVPSEYLTNIHIRDKLAAIKLGRYGEDLLFYLYY  SEQ ID NO-3
239 AHELDVKVPEEYMTHNHIRDKLPPLRLNKVSEDVLFYLFY  SEQ ID NO-8
83  -GVLQSTVTPPCFNSIQNDQQ----RVALFQDETLFFLFY  SEQ ID NO-9

190 MNGGDVLQLLAAVELFNRDWRYHKEERVWITRAPGMEPTM  SEQ ID NO-3
279 NCPNEIYQVAAACELYAREWRFHKSEQVWLTRSQYGGVKE  SEQ ID NO-8
118 KHPGTVIQELTYLELRKRNWRYHKTLKAWLTKDPMMEPIV  SEQ ID NO-9

230 KTNTY-ERGTYYFFDCLXWRKVAKEFHLEYDKLEERPHLP  SEQ ID NO-3
319 QTGNY-EKGHYNVFDQMQWRKIPKELKLEYKELEDRPKMP  SEQ ID NO-8
158 SADGLSERGSYVFFDPQRWEKCQRDFLL------------  SEQ ID NO-9

269 --------STFNY---------------------------  SEQ ID NO-3
358 QSVSGQPTPYKYFFQGPQFPSGPETGLMLQMHNLTLGTGG  SEQ ID NO-8
186 --------------FY------------------------  SEQ ID NO-9

274 --------NPAQQAF                           SEQ ID NO-3
398 GGGGQITPPAPAGLNGVMGGGGVGAAGIGGINVQPGAVPS  SEQ ID NO-8
188 ---------------NAIM                       SEQ ID NO-9

280                                           SEQ ID NO-3
438 AARATPN                                   SEQ ID NO-8
191                                           SEQ ID NO-9
```

FIGURE 6B

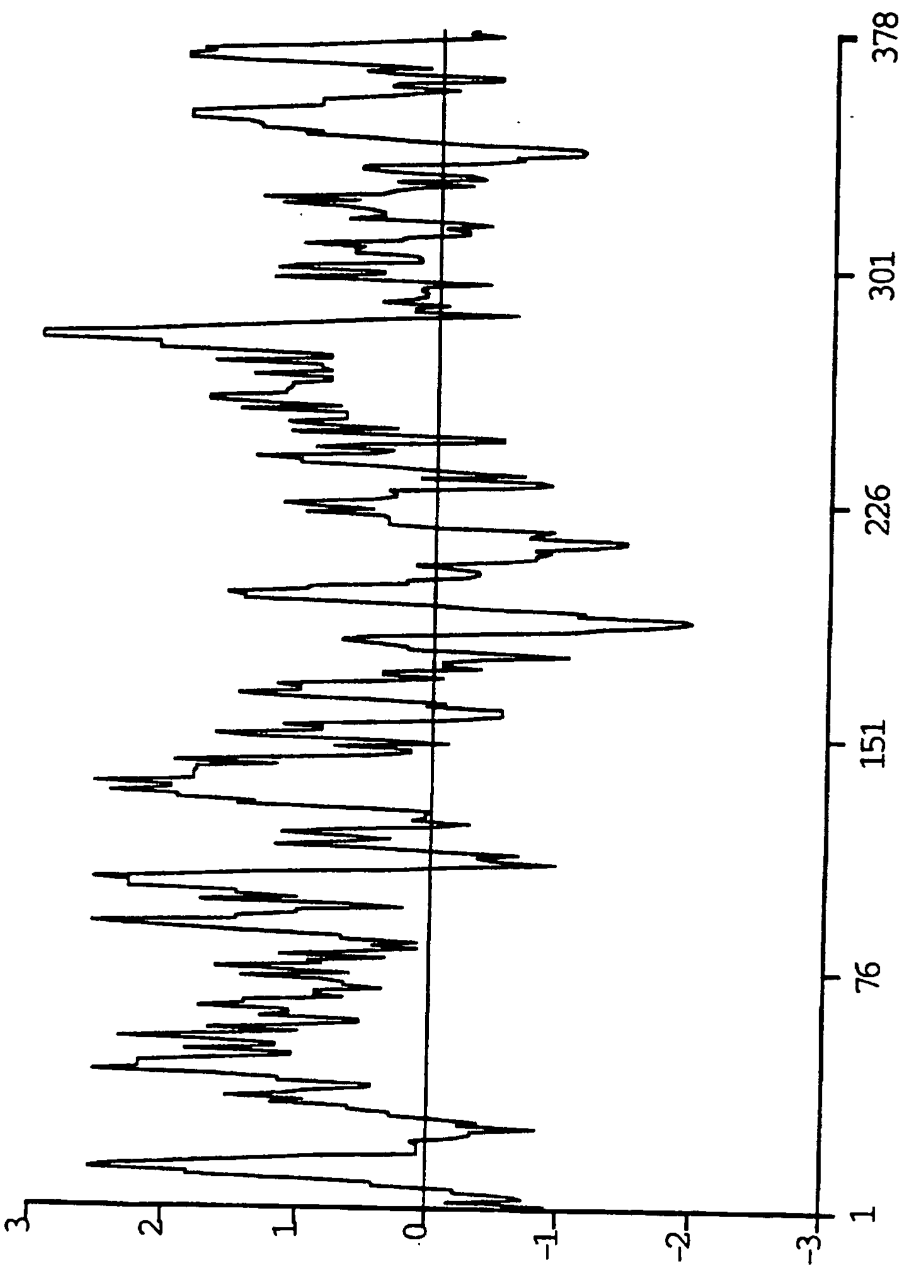


FIGURE 7

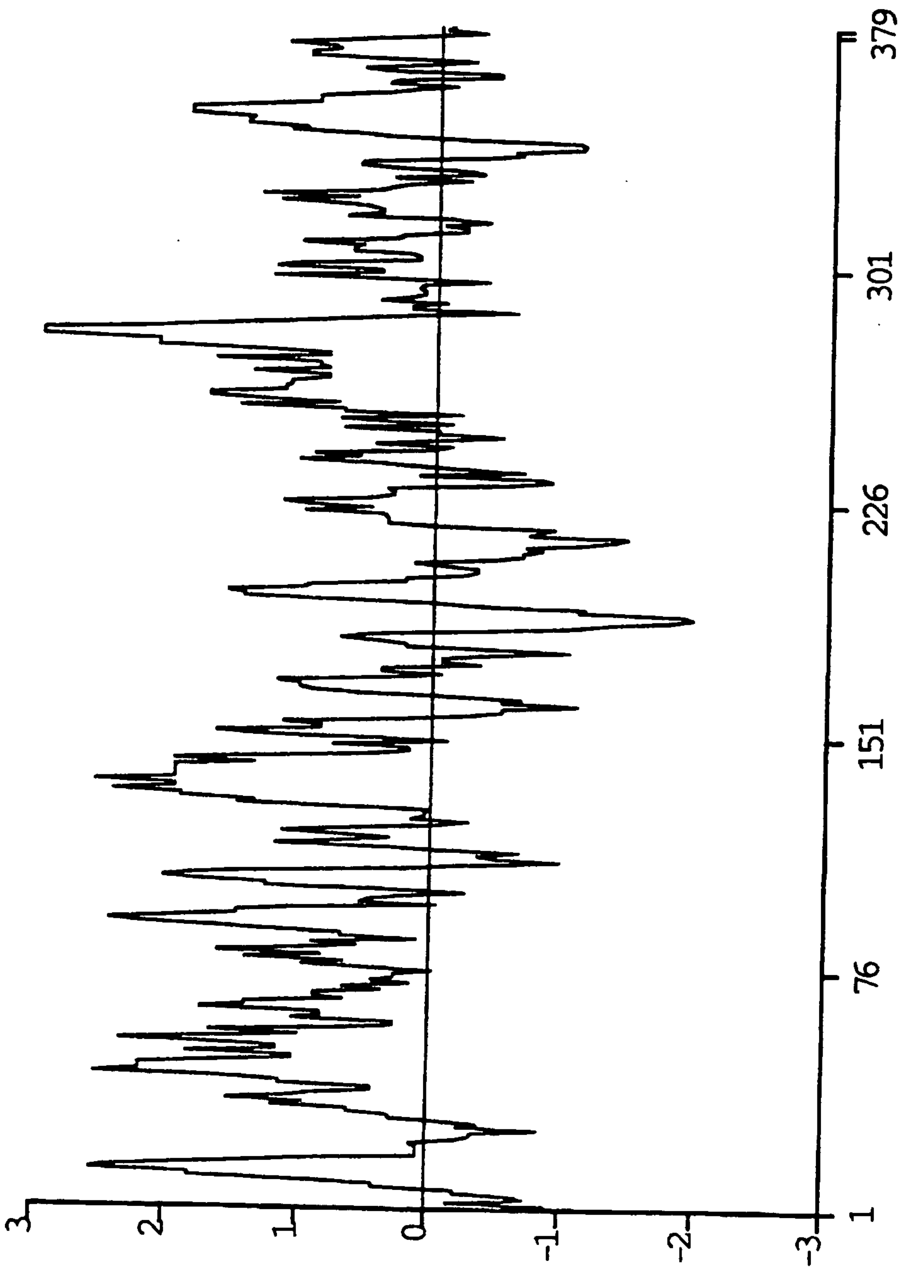


FIGURE 8

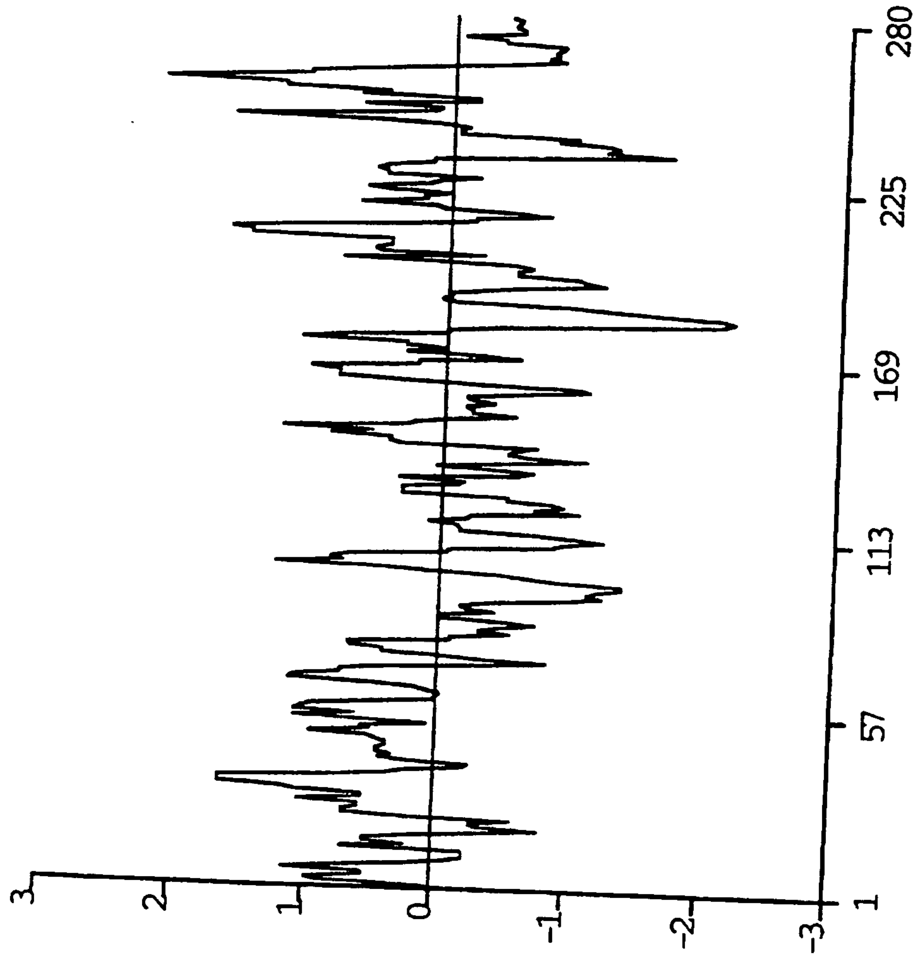


FIGURE 9

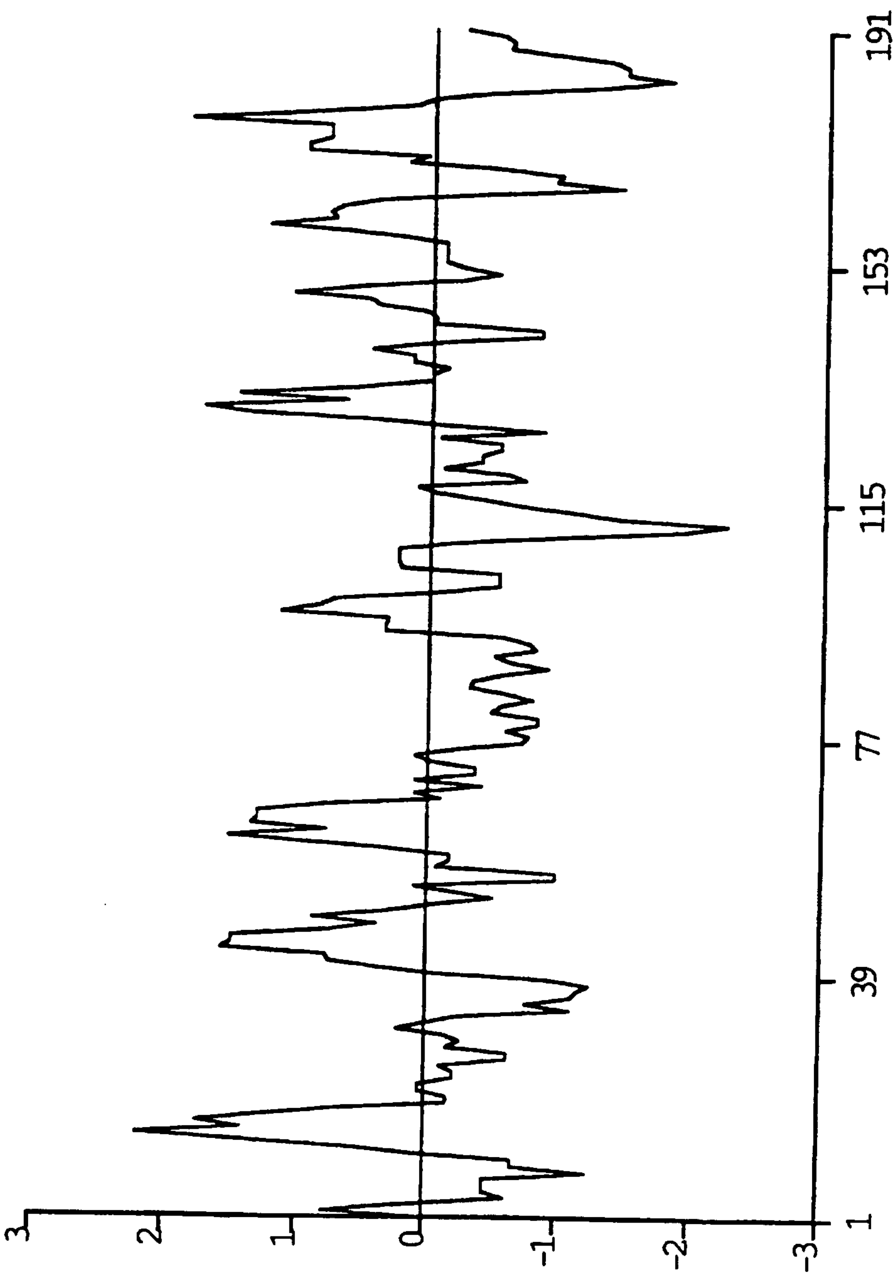


FIGURE 10

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 1/21, C07K 14/47, 16/18, A61K 38/17

A3

(11) International Publication Number: WO 98/11220

(43) International Publication Date: 19 March 1998 (19.03.98)

(21) International Application Number: PCT/US97/16174

(22) International Filing Date: 12 September 1997 (12.09.97)

(30) Priority Data:
08/712,708 12 September 1996 (12.09.96) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/712,708 (CIP)
Filed on 12 September 1996 (12.09.96)

(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). ZWEIGER, Gary, B. [US/US]; 513 S. Fremont Street, San Mateo, CA 94402 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(88) Date of publication of the international search report: 16 July 1998 (16.07.98)

(54) Title: NOVEL HUMAN CELL DIVISION CYCLE PROTEINS

| Library | Lib Description | Abun | Pct Abun |
|---|---|---|---|
| FIBRNGT01 | GD23A fibroblasts, radiation 5 min | 1 | 0.1664 |
| PITUNOR01 | pituitary, 16-70 M/F, RP | 1 | 0.1233 |
| MYOMNOT01 | uterus, myometrium, 43 F | 1 | 0.0409 |
| STOMTUT01 | stomach tumor, 52 M, match to STOMNOT02 | 1 | 0.0367 |
| BRAITUT02 | brain tumor, metastasis, 58 M | 2 | 0.0338 |
| STOMNOT02 | stomach, 52 M, match to STOMTUT01 | 1 | 0.0308 |
| LUNGNOT09 | lung, fetal M | 1 | 0.0286 |
| PTHYTUM01 | parathyroid tumor, adenoma, M/F, NORM, WM | 1 | 0.0278 |
| LNODNOT03 | lymph node, 67 M | 1 | 0.0265 |
| BRAITUT13 | brain tumor, meningioma, 68 M | 1 | 0.0262 |
| DUODNOT02 | small intestine, duodenum, 8 F | 1 | 0.0262 |
| BRAINOT03 | brain, 26 M | 1 | 0.0185 |
| HNT2RAT01 | hNT-2 cell line, teratocarcinoma, treated RA | 1 | 0.0185 |
| LUNGNOT04 | lung, 2 M | 1 | 0.0183 |
| UTRSNOT02 | uterus, 34 F | 1 | 0.0166 |
| NGANNOT01 | ganglioneuroma, 9 M | 1 | 0.0155 |
| BRAINOM01 | brain, infant F, NORM, WM | 3 | 0.0134 |
| UCMCL5T01 | mononuclear cells, treated IL-5 | 1 | 0.0125 |

(57) Abstract

The present invention provides novel human cell division cycle proteins (collectively called HCDC) and polynucleotides which identify and encode HCDC. The invention also provides genetically engineered expression vectors and host cells comprising the nucleic acid sequences encoding HCDC. The invention also provides pharmaceutical compositions containing HCDC or antagonists to HCDC, and in the use of these compositions for the treatment of diseases associated with the expression of HCDC. Additionally, the invention provides for the use of antisense molecules to polynucleotides encoding HCDC for the treatment of diseases associated with the expression of HCDC. The invention also provides diagnostic assays which utilize the polynucleotide, or fragments or the complement thereof, to hybridize to the genomic sequence or transcripts of polynucleotides encoding HCDC or anti-HCDC antibodies which specifically bind to HCDC.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# INTERNATIONAL SEARCH REPORT

Intern. ial Application No
PCT/US 97/16174

**A. CLASSIFICATI N OF SUBJECT MATTER**
IPC 6 C12N15/12 C12N1/21 C07K14/47 C07K16/18 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | K. DAI ET AL.: "Physical interaction of mammalian Cdc37 with CDK4" J. BIOL. CHEM., vol. 271, no. 36, 6 September 1996, AM. SOC. BIOCHEM. MOL.BIOL.,INC.,BALTIMORE,US, pages 22030-22034, XP002051554 see the whole document --- | 1-12 |
| X | L. STEPANOVA ET AL.: "Mammalian p50Cdc37 is a protein kinase-targeting subunit of Hsp90 that binds and stabilizes Cdk4" GENES & DEVELOPMENT, vol. 10, no. 12, 15 June 1996, CSH LABORATORY PRESS, NEW YORK,US, pages 1491-1502, XP002051555 cited in the application see the whole document --- | 1-12 |

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 9 January 1998 | 27-05-1998 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | HORNIG H. |

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | L. STEPANOVA ET AL.: "Mammalian p50Cdc37 is a protein kinase targeting subunit of Hsp90 that bind and stabilizes CDK4" EMBL SEQUENCE DATABASE, 16 June 1996, HEIDELBERG, BRD, XP002051556 Accession no. U43077 --- | 1-12 |
| X | WO 95 33819 A (MITOTIX INC) 14 December 1995 SEQ ID nos. 21,45 --- | 1-12 |
| A | N. GRAMMATIKAKIS ET AL.: "A novel glycosaminoglycan-binding protein is the vertebrate homologue of the cell cycle control protein, Cdc37" J. BIOL. CHEM., vol. 270, no. 27, 7 July 1995, AM. SOC. BIOCHEM. MOL.BIOL.,INC.,BALTIMORE,US, XP002051557 cited in the application see the whole document ----- | 1-12 |

1

**INTERNATIONAL SEARCH REPORT**

In. .ational application No.
PCT/US 97/ 16174

**Box I Observati ns where certain claims were found unsearchable (Continuati n of it m 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
See annex

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

See annex

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-12

**Remark on Protest**

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATI N CONTINUED FR M PCT/ISA/ 210

1. Claims: 1-12

A substantially purified human cell division cycle protein comprising the amino acid sequence of SEQ ID no.1 or fragments thereof; an isolated purified polynucleotide sequence encoding said protein, whereby the sequence is SEQ ID no.2 or variants thereof; polynucleotide sequence which is complementary to said sequence or variants thereof; an expression vector containing SEQ ID no.2; a host cell comprising said vector; a method for producing said polypeptide of SEQ ID no.1; a pharmaceutical composition comprising said polypeptide; a purified body which binds to said polypeptide; a purified antagonist which specifically regulated or modulates the activity of said polypeptide having the SEQ ID no.1; a method for treating cancer comprising administering to a subject in need of such treatment an effective amount of said pharmaceutical.

2. Claims: 13-24

Idem as subject 1 but limited to SEQ ID nos. 3 and 4.

Remark : Although claim 12 is directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the compound/composition.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern. nal Application No
PCT/US 97/16174

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 9533819 A | 14-12-95 | US 5691147 A | 25-11-97 |
| | | AU 2662795 A | 04-01-96 |